- 1 -

DETECTION of EPIGENTIC ABNORMALITIES and DIAGNOSTIC METHOD BASED THEREON

The present invention relates to identification of epigenetic abnormalities.

5   More particularly, the present invention relates to diagnosis of diseases based on DNA methylation differences, and identification and isolation of genes that cause such diseases.

BACKGROUND OF THE INVENTION

10

Substantial progress has been made in recent years with respect to the diagnosis and treatment of diseases in which a single defective gene is responsible. Traditional linkage studies have effectively isolated the causal gene and allowed for the further development of diagnostic tests and furthered research into treatments such

15   as gene therapy for conditions such as cystic fibrosis, Duchennes muscular dystrophy, Huntington's disease and fragile X syndrome. However, similar progress has not been made in diseases caused by mutations in multiple genes. Traditional linkage studies in complex diseases such as schizophrenia, bipolar disorder, cancers and diabetes have only succeeded in isolating chromosome regions, often containing 200-300

20   genes. The ability to screen such a large number of genes is clearly a time-consuming and daunting task.

Epigenetic mechanisms can be an important factor in complex, multi-factorial diseases such as cancers. Epigenetics refers to modifications in gene expression that

25   are brought about by heritable, but potentially reversible changes in DNA methylation and chromatin structure (Henikoff S, Matzke MA Exploring and explaining epigenetic effects. Trends Genet 1997,13(8):293-5; Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. Nat Genet 1999, 22(2):203-206; Gonzalgo, M.L. and Jones, P.A. (1997) Mutagenic

30   and epigenetic effects of DNA methylation. Mutat. Res. 386(2), 107-18; Razin, A. and Shemer, R. (1999) Epigenetic control of gene expression. Results Probl. Cell. Differ. 25, 189-204; Lyko, F. and Paro, R. (1999) Chromosomal elements conferring epigenetic inheritance. Bioessays 21(10), 824-32). DNA methylation of the binding

- 2 -

sites for transcription factors changes the affinity of such factors for regulatory sequences, which affects the transcriptional activity of a gene (Ehrlich M and Ehrlich K (1993) Effect of DNA methylation and the binding of vertebrate and plant proteins to DNA. In: Jost JP and Saluz P (eds) DNA Methylation: Molecular Biology and Biological Significance pp. 145-168. Birkhauser Verlag, Basel, Switzerland; Riggs A, Xiong Z, Wang L, and LeBon JM (1998) Methylation dynamics, epigenetic fidelity and X chromosome structure. In: Wolffe AP (ed) Epigenetics, pp. 214-227. John Wiley & Sons, Chistester). In addition to positional effects of methylated cytosines, density in a gene regulatory region also contributes to gene activity. This type of regulation is mediated by methylated cytosine binding proteins and acetylation of histones (Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, and Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics 19: 187-91; Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, and Bird A (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-9; Robertson KD and Wolffe AP (2000) DNA methylation in health and disease. Nature Review Genet 1:11-9).

Methylation can occur within cytosine-guanosine islands (CpG islands) that are typically between 0.2 to about 1 kb in length and are located upstream of many housekeeping and tissue-specific genes, but may also extend into protein coding regions. Methylation of cytosine residues contained within CpG islands of certain genes has been inversely correlated with gene activity. This could lead to decreased gene expression by a variety of mechanisms including, for example, disruption of local chromatin structure, inhibition of transcription factor-DNA binding, or by recruitment of proteins which interact specifically with methylated sequences indirectly preventing transcription factor binding. Some studies have demonstrated an inverse correlation between methylation of CpG islands and gene expression. Tissue-specific genes are usually unmethylated within the receptive target organ cells but are methylated in the germline and in non-expressing adult tissues. CpG islands of constitutively-expressed housekeeping genes are normally unmethylated in the germline and in somatic tissues.

- 3 -

In comparison to the role of DNA hypermethylation in disease, the role of DNA hypomethylation has attracted much less attention from researchers. However, DNA hypomethylation has been generally linked to disease states. For example, cancerous tissue has been shown to have lower levels of DNA methylation when

5   compared to normal tissue (Lapeyre, J. N. and Becker, F. F. (1979). 5-Methylcytosine content of nuclear DNA during chemical hepatocarcinogenesis and in carcinomas which result. Biochem Biophys Res Commun 87, 698-705; Gama-Sosa, M. A., Slagel, V. A., Trewyn, R. W., Oxenhandler, R., Kuo, K. C., Gehrke, C. W., and Ehrlich, M. (1983). The 5-methylcytosine content of DNA from human tumors.

10   Nucleic Acids Res 11, 6883-94; Feinberg, A. P., Gehrke, C. W., Kuo, K. C., and Ehrlich, M. (1988). Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res 48, 1159-61). Furthermore, activation of oncogenes as a result of DNA hypomethylation has been proposed (Feinberg, A. P. and Vogelstein, B. (1983) Hypomethylation of ras oncogenes in primary human cancers. Biochem

15   Biophys Res Commun 111, 47-54). Although a significant correlation between DNA hypomethylation and diseased states has been established, there is a need for methodology for identifying specific DNA hypomethylation-based epigenetic abnormalities that may increase the risk of developing a diseased state.

20   US5871917 discloses methods for detecting epigenetic abnormalities comprising: restriction of genomic DNA with a methylation-sensitive restriction enzyme (a restriction enzyme that cleaves an unmethylated site, but does not cleave the same site if it is methylated) that leaves an overhang; ligation of adaptors to the overhangs; PCR amplification with primers directed to the adaptors; followed by a

25   subtractive hybridization to eliminate house keeping genes; and a second round of PCR amplification with a second set of primers directed to a second set of adaptors. A problem with this design is that the method is limited to a restriction enzyme that leaves overhangs and, further, the method is complicated due to the ligation of two sets of adaptors.

30

    WO99/01580 discloses methods for detection of genomic imprinting disorders based on digestion of genomic DNA with methylation-sensitive restriction enzymes and PCR amplification using primers. One embodiment, directed to the detection of

unmethylated sequences, requires the use of a restriction enzyme that leaves overhangs and the use of exogenous adaptors, and therefore suffers from similar disadvantages as those described above in regards to US5871917. Another embodiment, directed to the detection of methylated sequences, uses primers directed

5   to endogenous elements such that exogenous adaptors are not required, but these primers are required to be positioned on either side of a methylation-sensitive restriction site. Since a methylation sensitive restriction enzyme will cut an unmethylated site, this method can only be used to amplify the methylated sequences, and cannot produce an unmethylated sequence which will be cut in between the two

10 primers.

It is an object of the present invention to overcome disadvantages of the prior art.

15 The above object is met by a combination of the features of the main claims. The sub claims disclose further advantageous embodiments of the invention.

SUMMARY OF THE INVENTION

20 The present invention relates to detection of epigenetic abnormalities and diagnosis of diseases associated with epigenetic abnormalities, and identification and isolation of genes that cause such diseases.

25 According to the present invention there is provided a method of detecting an epigenetic abnormality associated with a disease comprising:
identifying, within a eukaryotic genome, a locus having a hypomethylated sequence specific for said disease and an endogenous multi-copy DNA element. The method can comprise separate steps of identifying a disease-specific hypomethylated

30 sequence and identifying an endogenous multi-copy DNA element, where the steps may be performed in any order, so long as a locus is identified that has both a disease-specific hypomethylated sequence and an endogenous multi-copy DNA element. The disease-specific hypomethylated sequence and the endogenous multi-copy DNA

- 5 -

element will often be within 20 kilobases of separation, for example, within 20, 10, 5, 2, 1, 0.1 kilobases of each other, or may even be so close as to overlap. The endogenous multi-copy DNA element can include any retroelement that is normally methylated examples of which include, without limitation, endogenous retroviral sequences (ERV), Alu sequences, and LINE sequences. The endogenous multi-copy DNA element may be located within any eukaryotic genome including fungi, plants, and animals, with mammalian and human genomes being non-limiting examples of animal genomes.

In another aspect, the present invention provides a method of identifying a chromosomal region associated with a diseased state comprising:
identifying a locus, within DNA obtained from a diseased sample, that has a DNA sequence that is hypomethylated and an endogenous multi-copy DNA element, wherein the DNA sequence is methylated in a non-disease sample and wherein the chromosomal region consists of from about 1 to about 10 DNA coding sequences that are proximal to the identified locus. In a further aspect, a DNA coding sequence having an epigenetically altered expression pattern that contributes to a disease in an organism can be identified by comparing expression patterns of the DNA coding sequence located proximal to the disease-specific hypomethylated locus within a test sample that exhibits characteristics of said disease with expression patterns of a corresponding DNA coding sequence within a control sample to identify the DNA coding sequence having an epigenetically altered expression pattern. The DNA coding sequence may encode an RNA that remains non-translated, or may encode an RNA that is translated, at least partially, into a polypeptide.

In another aspect, the present invention provides a method of diagnosing an epigenetic abnormality correlated with a disease comprising:
identifying a DNA sequence that is hypomethylated within a locus that has an endogenous multi-copy DNA element and is obtained from a diseased sample, wherein the DNA sequence is methylated in a non-disease sample.

- 6 -

According to yet another aspect of the present invention there is provided a method of detecting an epigenetic abnormality associated with a disease, the method comprising:

a) extraction of genomic DNA from a sample that exhibits characteristics of a disease;

b) digestion of the genomic DNA with a methylation-sensitive restriction enzyme to produce a pool of restricted DNA fragments;

c) fractionation of the pool of restricted DNA fragments to obtain DNA fragments of a desired size;

d) amplification of at least a segment of the DNA fragments of a desired size with primers that anneal to an endogenous DNA element to produce a PCR product;

e) cloning of the PCR product into a sequencing vector;

f) sequence determination of the PCR product to obtain a sequence of the PCR product;

g) comparing the sequence against a genomic database to assign a locus for the epigenetic abnormality associated with a disease.

The sample from which DNA is extracted may be any cell, tissue, organ or other suitable specimen that exhibits characteristics of a disease. For example, without wishing to be limiting, in an individual suffering from schizophrenia, Huntingdon's disease, or bipolar disorder a sample may be obtained from brain tissue.

Any endogenous multi-copy DNA element that is found to have epigenetic abnormalities associated with a disease can be PCR amplified according to the present invention. In a further aspect, the endogenous DNA element is a multi-copy DNA element. In a still further aspect, the multi-copy DNA element is selected from the group consisting of LINE, SINE, L1, and Alu.

In still another aspect, the present invention provides a method of identifying a gene having an epigenetically altered expression pattern that contributes to a disease in an organism, the method comprising:

a) extraction of genomic DNA from a sample that exhibits characteristics of a disease;

b) digestion of the genomic DNA with a methylation-sensitive restriction enzyme to produce a pool of restricted DNA fragments;

c) fractionation of the pool of restricted DNA fragments to obtain DNA fragments of a desired size;

d) amplification of at least a segment of the DNA fragments of a desired size with primers that anneal to an endogenous DNA element to produce a PCR product;

e) cloning of the PCR product into a sequencing vector;

f) sequence determination of the PCR product to obtain a sequence of the PCR product;

g) comparing the sequence against a genomic database to assign a locus for said epigenetic abnormality associated with a disease;

h) searching said database to identify a gene located proximal to said locus;

i) comparing expression patterns of said gene located proximal to said locus within a test sample that exhibits characteristics of said disease with expression patterns of a corresponding gene within a control sample to identify said gene having an epigenetically altered expression pattern.

Genes can be identified in accordance with the present invention from any eukaryotic organism including, plants and animals, where epigenetic abnormality is associated with the occurrence of disease.

In yet another aspect, the present invention provides a method of isolating a probe for detecting an epigenetic abnormality associated with a disease in an animal, said method comprising:

a) extraction of genomic DNA from a sample that exhibits characteristics of said disease;

b) digestion of said genomic DNA with a methylation-sensitive restriction enzyme to produce a pool of restricted DNA fragments;

c) fractionation of said pool of restricted DNA fragments to obtain DNA fragments of a desired size;

d) amplification of at least a segment of said DNA fragments of a desired size with primers that anneal to an endogenous DNA element to produce a PCR product;

- 8 -

f) using said PCR product as said probe to detect said epigenetic abnormality associated with said disease in another sample.

In still another aspect, there is provided methods for detecting disease or diagnosing disease. In an aspect the present invention provides a method of detecting a disease associated with an epigenetic abnormality comprising, identifying, within a eukaryotic genome, a locus having a hypomethylated sequence specific for the disease and an endogenous multi-copy DNA element. In another aspect the present invention provides a method of diagnosing a disease correlated with an epigenetic abnormality comprising identifying a DNA sequence that is hypomethylated within a locus that has an endogenous multi-copy DNA element and is obtained from a diseased sample, the DNA sequence being methylated in a non-disease sample.

The methods of the present invention can be applied to any disease that occurs as a result of hypomethylation within a locus having an endogenous multi-copy DNA element, including Mendelian and non-Mendelian disease. Illustrative examples of diseases include, without limitation, Huntington's disease, schizophrenia, bipolar disorder, cancers, neuropsychiatric diseases, and diabetes.

This summary does not necessarily describe all necessary features of the invention but that the invention may also reside in a sub-combination of the described features.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features of the invention will become more apparent from the following description in which reference is made to the appended drawings wherein:

5

FIGURE 1 shows the localization of the cloned Alu elements.

FIGURE 2 shows DNA coding sequences that comprise or are located within very close proximity (within 100,000 bp) of cloned Alu elements.

10

FIGURE 3 shows sequences of cloned Alu elements in Example 4 (SEQ ID NO:29-263).

FIGURE 4 shows an alignment of a portion of cloned Alu elements in Example 1

15   (SEQ ID NO:6-28). Alignment file of cloned Alu sequences was created using CLUSTAL W Multiple Sequencing Alignment Program (http://clustal w.genome.ad.jp/).

- 10 -

DESCRIPTION OF PREFERRED EMBODIMENT

The invention relates to methods and compositions for identification of epigenetic abnormalities. More particularly, the present invention relates to diagnosis of diseases based on DNA methylation differences and identification of genes that cause such diseases. The present invention provides methods and compositions for detecting and isolating DNA sequences which are abnormally or differentially methylated in a diseased cell type when compared to a normal cell type.

Traditional linkage studies in complex diseases such as schizophrenia, bipolar disorder, cancers and diabetes have only succeeded in isolating chromosome regions, often containing 200-300 genes. The ability to screen such a large number of genes is clearly a time-consuming and daunting task. The present invention provides a short-cut in determining which genes within a 200-300 gene region are in fact responsible for the onset of a major disease such as diabetes, schizophrenia, cancers, or bipolar disorder. According to the present invention differentially modified, endogenous multi-copy DNA elements can act as markers for genes which are dys-regulated. Epigenetic analysis of so called "junk" DNA leads to a 'short-cut' in identification of specific genes, dys-regulation of which increases the risk to major disease.

The following description is of a preferred embodiment by way of example only and without limitation to the combination of features necessary for carrying the invention into effect.

The methylation patterns of DNA from tumor cells are generally different than those of normal cells (Laird et al., DNA Methylation and Cancer, 3 Human Molecular Genetics 1487, 1488 (1994)). Tumor cell DNA is generally undermethylated relative to normal cell DNA, but selected regions of the tumor cell genome may be more highly methylated than the same regions of a normal cell's genome. Hence, detection of altered methylation patterns in the DNA of a tissue sample is an indication that the tissue is cancerous. For example, the gene for

- 11 -

Insulin-Like Growth Factor 2 (IGF2) is hypomethylated in a number of cancerous tissues, such as Wilm's Tumors, rhabdomyosarcoma, lung cancer and hepatoblastomas (Rainner et al. 362 Nature 747-49 (1993); Ogawa, et al., 362 Nature 749-51 (1993); S. Zhan et al., 94 J. Clin. Invest. 445-48 (1994); P. V. Pedone et al., 3

5   Hum. Mol. Genet. 1117-21 (1994); H. Suzuki et al., 7
Nature Genet 432-38 (1994); S. Rainier et al., 55 Cancer Res. 1836-38 (1995)).

        Alteration of methylation may be a key, and common event, in the development of neoplasia and may play at least two roles in tumorigenesis:

10   1) DNA hypomethylation may cause an increase in proto-oncogene expression or DNA hypermethylation may decrease expression of a tumor supressor which contributes to neoplastic growth; and

            2) DNA hypomethylation may change chromatin structure, and induce abnormalities in chromosome pairing and disjunction. Such structural abnormalities

15   may result in genomic lesions, such as chromosome deletions, amplifications, inversions, mutations, and translocations, all of which are found in human genetic diseases and cancer.

        While the present invention can be used for detecting any alteration in

20   methylation, the present invention is particularly useful for detecting and isolating DNA fragments that are normally methylated but which, for some reason, are non-methylated in a proportion of cells. Such DNA fragments may normally be methylated for a number of reasons. For example, such DNA fragments may be normally methylated because they contain, or are associated with, genes that are

25   rarely expressed, genes that are expressed only during early development, genes that are expressed in only certain cell-types, and the like.

        As used herein, hypomethylation means that at least one cytosine in a CG or CNG di- or tri-nucleotide site in genomic DNA of a given cell-type does not contain

30   CH_3 at the fifth position of the cytosine base. Cell types that may have hypomethylated CGs or CNGs, such as, without limitation, CCGs, include any cell type that may be expressing a non-housekeeping function. This includes both normal cells that express tissue-specific or cell-type specific genetic functions, as well as

- 12 -

tumorous, cancerous, and similar cell types. Cancerous cell types and conditions which can be analyzed, diagnosed or used to obtaining probes by the present methods include, but are not limited to, Wilm's cancer, breast cancer, ovarian cancer, colon cancer, kidney cell cancer, liver cell cancer, lung cancer, leukemia,

5   rhabdomyosarcoma, sarcoma, and hepatoblastoma.

       A method of the present invention is directed to detection of an epigenetic abnormality comprising identifying, within a eukaryotic genome, a locus having a hypomethylated sequence and an endogenous multi-copy DNA element. The method

10   can comprise separate steps of identifying a hypomethylated sequence and identifying an endogenous multi-copy DNA element, where the steps may be performed in any order, so long as a locus is identified that has both a hypomethylated sequence and an endogenous multi-copy DNA element. The hypomethylated sequence and the endogenous multi-copy DNA element will often be within 20 kilobases of separation,

15   for example, within 20, 10, 5, 2, 1, 0.1 kilobases of each other, or may even be so close as to overlap. The endogenous multi-copy DNA element can include any retroelement, examples of which include, without limitation, endogenous retroviral sequences (ERV), Alu sequences, L1 sequences, SINE sequence, and LINE sequences. The endogenous multi-copy DNA element will be located within any

20   eukaryotic genome including fungi, plants, and animals, with mammalian and human genomes being non-limiting examples of animal genomes.

       Without wishing to be bound by theory, hypermethylation in a locus having a retroelement, within eukaryotic genomes, can function to suppress transcriptional

25   activity of the retroelement. Hypomethylation may underlie disease by undesired removal of the suppression of transcriptional activation of a retroelement and/or surrounding genes. As such the combination of a hypomethylated sequence and a retroelement can serve as a useful marker for an aberrant regulation of DNA sequence expression that can be a factor in a diseased state.

30

       As will be recognized by persons skilled in the art, various techniques may be used to identify a locus having a hypomethylated sequence and an endogenous multi-

- 13 -

copy DNA element. For example, techniques that are known to be reliable for
detecting differences in DNA methylation include, but are not limited to:

   - methylation-sensitive restriction enzymes (Issa J.P., et al. (1994) Nature
Genetics 7:536-40);

5   - methylation-sensitive arbitrarily primed PCR (Liang G, et al. (2002)
Identification of DNA methylation differences during tumorigenesis by methylation-
sensitive arbitrarily primed polymerase chain reaction. Methods 27(2):150-5);

   - sequencing of sodium bisulfite-induced modifications of genomic DNA
(Frommer M, et al. (1992) A genomic sequencing protocol that yields a positive
10   display of 5-methylcytosine residues in individual DNA strands);

   - methylation-specific PCR based on differential hybridization of PCR primer
to DNA initially modified by bisulfite treatment (Herman JG, et al. (1996)
Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands.
Proc Natl Acad Sci USA 93:9821-26; Fan X, et al. (Improvement of the methylation
15   specific PCR technical conditions for the detection of p16 promoter hypermethylation
in small amounts of tumor DNA. Oncology Rep 9:181-3); or

   - methylation-sensitive single nucleotide primer extension based on bisulfite-
modification of DNA followed by differential incorporation of labelled nucleotides to
a primer that is designed to hybridise immediately upstream of a methylation site
20   (Gonzalgo and Jones (1997) Rapid quantitation of methylation differences at specific
sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPe)
Nucleic Acids Research 25:2529-31).

      Several techniques are also available for identifying an endogenous multi-
25   copy DNA element within a locus. For example, endogenous multi-copy DNA
elements can be localized *in silico* for genomes that have been sequenced, annotated
and deposited within public, private, or commercial databases. As another example,
PCR primers can be used to detect the presence of an endogenous multi-copy DNA
element within a larger DNA sequence. As yet another example, Southern
30   hybridisation with probes comprising an endogenous multi-copy DNA element
sequence can be used for identifying and localizing the presence of the multi-copy
DNA element within a larger DNA sequence.

- 14 -

Hypomethylation of genomic sequences can be determined by using both methylation-sensitive restriction enzyme analysis, and genomic sequencing. Various restriction enzymes are available that digest demethylated sequences, while leaving methylated sequences intact. An advantage of methylation-sensitive restriction

5   enzyme analysis is that it produces DNA fragments that have 5' and 3' ends that were demethylated at the time of digestion. As a result it is a quick method of localizing demethylated sequences within a particular restriction sequence within a larger DNA sequence, such as a locus, chromosome, or even a whole genome. Methylation-sensitive restriction enzyme analysis, as well as examples of various methylation-

10   sensitive restriction enzymes, are described in greater detail below.

Methylation-sensitive DNA sequencing, while not as quick a method as restriction enzyme analysis, can provide specific sequence information with regards to any methylation site, regardless of its inclusion within a restriction enzyme site.

15   Maxam and Gilbert chemical cleavage sequencing protocols have been modified and developed to determine methylation status of sequences within a gene, with the absence of a band in all tracks of a sequencing gel indicating the presence of a 5-methylcytosine residue (Church and Gilbert (1984) Proc Natl Acad Sci USA 81:1991-95; Saluz and Jost (1989) Proc Natl Acad Sci USA 86:2602-6; Pfeifer GP, et al.

20   (1989) Science 246:810-13).

Another method of methylation-sensitive DNA sequencing involves exposing genomic DNA to sodium bisulfite (Frommer M, et al. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA

25   strands) under conditions where cytosine residues are converted to uracil residues, while 5-methylcytosine residues remain nonreactive. One or both strands of the bisulfite-modified genomic DNA can then be PCR amplified using pairs of strand specific primers. As the bisulfite reaction protocol produces single DNA strands that can no longer achieve 100% complementary basepairing (for example reacting double

30   stranded DNA consisting of 5'-TCTC-3' base paired to 5'-GAGA-3' with sodium bisulfite yields single strands of 5'-TUTU-3' and 5'-GAGA-3' such that 100% complementary base pairing can no longer be achieved), pairs of PCR primers can be designed such that they anneal in a strand-specific fashion and produce PCR products

- 15 -

for each of the single bisulfite-modified DNA strands. The PCR products can then be subject to any combination of assays available to skilled persons including, without limitation, sequencing, cloning, methylation-specific PCR, Ms-SNuPe, or microarrays. Bisulfite-modified DNA templates can be conveniently produced using

5   the EZ DNA methylation Kit™ developed by Zymo Research.

The combination of methylation-specific technology and array technology may be particularly useful for high throughput applications. For example, fragments of bisulfite-modified DNA could be analysed using microarrays having probes that

10   were specific for identified hypomethylated sequences. As another example, an array of primers could be developed for analysing each potential demethylation site by Ms-SNuPe assay within a DNA sequence, such as a locus, chromosome, or even a whole genome.

15   The above techniques can also be used in diagnosis of disease. For example, once one or more than one hypomethylated sequence have been correlated with a disease state, DNA obtained from a subject having the disease can be treated with sodium bisulfite, followed by Ms-SNuPe or methylation-specific PCR using primers that are specific for the correlated hypomethylated sequence(s). As another example,

20   diagnosis of disease can be achieved by digesting DNA, from a diseased sample, with a methylation-sensitive restriction enzyme that yields a different size fragment when digesting DNA from a diseased sample compared to DNA obtained from a normal sample; determination of the disease-specific restriction fragment size can be achieved through any standard method including, Southern analysis.

25

It will be understood that diagnostic methods of the present invention may be used to identify the presence of a disease in a subject, or may be used to identify a predisposition of a subject to develop a disease. As such the diagnostic methods of the present invention encompass pre-diagnosis of disease.

30

Accordingly, the present invention is directed to a method of diagnosing an epigenetic abnormality correlated with a disease comprising identifying a hypomethylated sequence within a locus that has an endogenous multi-copy DNA

- 16 -

element, wherein the hypomethyated sequence is methylated in a normal sample. The strength of correlation between the presence of a particular hypomethylated sequence and a disease may vary. The strength of correlation can be expressed in terms of percentage of true positives (the number of people who develop a disease divided by

5   the number of people who test positive). Example 2 shows a 100% correlation between Huntingdon's disease and the presence of a locus having a hypomethylated sequence and an Alu sequence (the Alu sequence being located ~4Kb downstream of the (CAG)n/(CTG)n repeat region of the HD gene). As such Huntingdon's disease is an example of a particularly successful use of the diagnostic methods of the present

10   invention. Furthermore, the diagnostic methods of the present invention can be successfully used in cases where strength of correlation between disease and hypomethylated sequence is lower than 100%, and could be as low as 50%, 40%, 30% or 20%, or even lower. The strength of correlation that is required for successful use of the diagnostic methods of the invention may depend on several factors that can

15   be ascertained by persons skilled in the art, one of these factors being the strength of correlation provided by diagnostic methods that are available in the marketplace. For example, in a disease where no diagnostic method is currently available the diagnostic methods of the present invention may be useful even if providing a strength of correlation that is lower than 20%. Persons skilled in the art will

20   recognize, that strength of correlation may include other factors in addition to the percentage of true positives, for example, a percentage of false positives (the number of people who do not develop a disease divided by the number of people who test positive). Again, as was the case for the desired percentage of true positives, the percentage of false positives that can be tolerated may depend on the number of false

25   positives being generated by commercially available diagnostic methods.

       Identification of hypomethylated sequences and endogenous multi-copy DNA elements can be accomplished using any suitable technique, or any other technique

30   that is convenient to the skilled technician. In order to illustrate the variability that can be incorporated in the present method for identifying a locus that has a hypomethylated sequence and a retroelement, for example, an Alu retroelement, the following non-limiting protocols are provided:

Protocol (A)

a) digest genomic DNA with a methylation-sensitive restriction enzyme (which digests hypomethylated sequences) to produce a pool of restricted DNA fragments,

b) fractionate the pool of restricted DNA fragments to obtain DNA fragments of a desired size,

c) amplify at least a segment of the DNA fragments of a desired size with primers that anneal to an Alu sequence to produce a PCR product having at least a portion of the Alu sequence,

d) determine the sequence the PCR product, and

e) compare said sequence against a genomic database to assign a locus for the PCR product having the at least a portion of the Alu sequence.

Protocol (B)

a) determine locations of Alu sequences *in silico* within a genomic database to obtain dataset of loci having Alu sequences,

b) modify genomic DNA from test and control samples by reacting with sodium bisulfite whereby cytosine is converted to uracil while 5-methylcytosine is unreacted,

c) amplify one or both strands of the converted DNA using pairs of strand-specific primers (primers are chosen such that they flank the Alu sequence at an appropriate distance, for example, 10 kilobases) to produce one (if only one strand amplified) or two (if both strands amplified) PCR products per loci under investigation,

d)(i) identify hypomethylated sequences by sequencing PCR products and identifying a C to T conversion in PCR product sequences derived from test samples compared to a lack of a C to T conversion in a corresponding nucleotide position in PCR product sequences derived from control samples; or

(ii) identify hypomethylated sequence by comparing test and control PCR products treated with restriction enzyme(s) that are appropriately chosen to distinguish between a methylated and bisulfite unreacted CG or CNG sequence versus a demethylated and bisulfite converted TG or TNG sequence (to obtain predicted

- 18 -

methylated and demethylated restriction maps any standard software can be used to convert all CG to XG then convert all C to T then convert all X to C and then produce a software predicted restriction map to obtain a methylated map, while conversion of all C to T followed by producing a software predicted restriction map provides a

5   demethylated map), or

(iii) identify hypomethylated sequence by comparing test and control PCR products in Ms-SNuPE assay (Gonzalgo and Jones (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPe) Nucleic Acids Research 25:2529-31) for each potential

10   demethylatation site (an advantage of this technique is that multiple methylation sites can be analysed in each by using a multiplex primer strategy with primers being designed to terminate immediately upstream of each methylation site in accordance with analysis of sequences flanking the identified Alu sequence), or

(iv) identify hypomethylated sequence by comparing the test and control PCR

15   products in methylation-specific PCR assays where primers are designed for differential primer annealing to an in silico predicted methylation site on the basis of bisulfite-induced C to T conversions;

Protocol (C)

20   a) determine locations of Alu sequences *in silico* within a genomic database to obtain dataset of loci having Alu sequences,

b) modify genomic DNA from test and control samples by reacting with sodium bisulfite whereby cytosine is converted to uracil while 5-methylcytosine is unreacted, and

25   c) identify hypomethylated sequence by comparing the test and control bisulfite-modified genomic DNA samples in methylation-specific PCR assays where primers are designed for differential primer annealing to an in silico predicted methylation site on the basis of bisulfite-induced C to T conversions;

30   Protocol (D)

a) identify locations of potential demethylation sites *in silico* within a genomic database to obtain dataset of loci having potential demethylation sites,

- 19 -

modify genomic DNA from test and control samples by reacting with sodium bisulfite whereby cytosine is converted to uracil while 5-methylcytosine is unreacted,

b) amplify bisulfite-converted DNA using strand-specific primers (primers are chosen such that they flank the potential demethylation site(s)) to produce PCR products,

c) identify hypomethylated sequence by comparing test and control PCR products in Ms-SNuPE assay for each potential demethylatation site to obtain an array of PCR products and loci having hypomethylated sequence(s),

d)(i) determine locations of Alu sequences in silico within dataset of loci having hypomethylated sequence(s), or

(ii) identify Alu sequences within the array of PCR products by any standard technique, for example, without limitation, Southern assay or PCR or DNA sequencing;

or,

Protocol (E)

a) identify locations of potential demethylation sites in silico within a genomic database to obtain dataset of loci having potential demethylation sites, modify genomic DNA from test and control samples by reacting with sodium bisulfite whereby cytosine is converted to uracil while 5-methylcytosine is unreacted,

b) amplify bisulfite-converted DNA using strand-specific primers (primers are chosen such that they flank the potential demethylation site(s)) to produce PCR products,

c) identify hypomethylated sequence by sequencing test and control PCR products and identifying a C to T conversion in PCR product sequences derived from test samples compared to a lack of a C to T conversion in a corresponding nucleotide position in PCR product sequences derived from control samples,

d) (i) determine locations of Alu sequences in silico within dataset of loci having hypomethylated sequence(s),

(ii) identify Alu sequences within the array of PCR products by any standard technique, for example, without limitation, Southern assay or PCR or DNA sequencing;

- 20 -

Any of the above protocols can be used to identify loci having a hypomethylated sequence and a multi-copy DNA element within a test sample compared to a control sample. Usually the test sample will be the genome of diseased tissue, while the control sample can be a corresponding tissue in a person not

5   suffering from the disease. However, persons skilled in the art will recognize other relevant test/control comparisons such as the control sample being any normal tissue from within a diseased animals own body (for example, cancerous liver tissue samples could be compared to non-cancerous liver tissue samples with both samples obtained from within the same subject). The methods of the present invention can be

10 applied to any disease that occurs as a result of hypomethylation within a locus having an endogenous multi-copy DNA element, including both Mendelian and non-Mendelian disease. Illustrative examples of diseases include, without limitation, cystic fibrosis, Duchennes muscular dystrophy, Huntington's disease, fragile X syndrome, schizophrenia, bipolar disorder, cancers and diabetes.

15

DNA analysed in accordance with methods of the present invention may be extracted from any sample that may have epigenetic abnormalities associated with a disease, for example, but not limited to cells of the following tissues: Epithelial Tissues, Exocrine Glands, Endocrine Glands, Connective Tissues, Adipose Tissue,

20 Cartilage, Bone, Blood, Muscle Tissues comprising Smooth, Skeletal or Cardiac Muscle Tissue, or Nervous Tissue comprising Brain Tissue. DNA can be extracted using standard techniques, known in the art, for isolating DNA from various samples such as cells , tissues, or organs, or other suitable specimens. Standard techniques for isolating DNA have are disclosed in reference textbooks or manuals such as

25 Sambrook, Fritsch, and Maniatis, Molecular Cloning: A Laboratory Manual (1989), Cold Spring Harbor.

The above-described non-limiting illustrative protocols specify the identification of Alu sequences. However, the methods of the invention are equally

30 applicable to other endogenous multi-copy DNA elements, for example, but not limited to, an L1 seqeunce, a SINE sequence, a LINE sequence, or an endogenous retroviral sequence (ERV).

- 21 -

A method of the present invention is directed to identifying a locus that has an increased probability of causing a diseased state comprising identifying a locus, within a genome obtained from a diseased sample, that has a hypomethylated sequence and an endogenous multi-copy DNA element, wherein the hypomethylated

5   sequence is methylated in a normal sample. An advantage of this method is that it provides a short cut for identification of causal factors of a disease, and further provides a short cut to identification of drug targets to treat disease. By concentrating on loci that have both a disease-specific hypomethylated sequence and an endogenous multi-copy DNA vast stretches of genomic DNA can be eliminated from analysis, and

10   analysis can be focused on DNA coding sequences that are proximal to, or comprise, the endogenous multi-copy DNA element and disease-specific hypomethylated sequence. For example, this assay may select from about 1 to about 10 DNA coding sequences from the disease-specific hypomethylated locus. By "DNA coding sequence" it is meant an open reading frame as commonly understood in the art

15

Techniques for analysing expression profiles of surrounding genes including, but not limited to, Northern, ELISA, reporter construct assays, microarray assay of RNA levels, dot blots, quantitative PCR, are well known to persons skilled in the art, and are not critical to the present invention. Any number of standard and available

20   techniques may be used to determine which of the genes proximal to a locus, identified in accordance with the present invention, are aberrantly regulated in a diseased state. The present invention provides for a quick way to focus available analytical resources on a set of about 1 to about 10 DNA coding sequences that are found to be surrounding or within a locus that has a disease-specific hypomethylated

25   sequence and an endogenous multi-copy DNA element. Usually, the dys-regulated gene which causes the diseased state will be found within the locus, or within a nucleotide sequence defined by the distance of about 1 to about 10 DNA coding sequences, and will be typically located within 1 to about 200 kilobases of the identified disease-specific hypomethylated locus. However, as seen in Table 3 this

30   separation may be less than 200 Kb and may vary, for example, without limitation, from about 100 Kb, to about 50 Kb, to about 5 Kb, to almost overlapping with the identified disease-specific hypomethylated locus.

- 22 -

By "dys-regulated gene" or "aberrantly regulated gene" it is meant a nucleotide sequence that is differentially regulated between a diseased and non-diseased sample.

5   The number of DNA coding sequences of less than about 10 compares favourably to a relatively larger range of 5 to 300 genes often contained within chromosomal regions identified by traditional genetic linkage studies. In a further aspect, a DNA coding sequence having an epigenetically altered expression pattern that contributes to a disease in an organism can be identified by comparing expression

10   patterns of the DNA coding sequence located proximal to the disease-specific hypomethylated locus within a test sample that exhibits characteristics of said disease with expression patterns of a corresponding DNA coding sequence within a control sample to identify the DNA coding sequence having an epigenetically altered expression pattern. The DNA coding sequence may encode an RNA that remains non-

15   translated, or may encode an RNA that is translated, at least partially, into a polypeptide.

A method of the present invention is directed to detection of epigenetic abnormalities associated with a non-Mendelian disease and comprises extraction of

20   genomic DNA from a non-Mendelian disease sample, such as diseased tissue or diseased population of cells; hydrolysis of this DNA with methylation-sensitive restriction enzymes, and subsequent fractionation of DNA fragments and purification of DNA fragments of a desired size, for example, but not limited to, shorter than 10 kB. These purified DNA fragments are further subjected to PCR amplification using

25   primers that hybridize to endogenous multi-copy DNA elements including, but not limited to, ALU or L1 elements. After that, PCR products of such elements are cloned and sequenced using standard molecular biology techniques known to the skilled artisan and the resultant sequences are mapped on the genome using any commercially or publicly available human genome database. These cloned multi-

30   copy elements indicate a loci of putative epigenetic abnormality or epigenetic dys-regulation and indicates genes that predispose a patient to a complex, non-Mendelian, multi-factorial disease, such as, but not limited to, cancers, diabetes, schizophrenia,

or bipolar disorder. Persons skilled in the art will recognize that this method can be used in regards to any disease, both non-Mendelian and Mendelian.

By the term "non-Mendelian disease" is meant any disease which etiologically
5   requires more than a single genetic abnormality. As such a non-Mendelian disease requires more than one factor, or in other words, is multi-factorial, and may comprise epigenetic alterations or abnormalities.

Epigenetics relates to higher order gene control mechanisms in eukaryotes that
10   activate or repress parts of the genome via changes in chromatin structure. These higher order gene control mechanisms form an important molecular basis of cell differentiation. Any changes in an organism brought about by alterations in the action of genes, where the changes do not require occurrence of any mutations, are called epigenetic changes. An epigenetic abnormality occurs when an epigenetic change
15   contributes or predisposes normal cells into becoming diseased cells. DNA methylation is an example of an epigenetic mechanism. The term DNA methylation refers to the addition of a methyl group to the cyclic carbon 5 of a cytosine nucleotide. A family of conserved DNA methyltransferases catalyzes this reaction. Normally, DNA methylation can be used, for example, but is not limited to, to methylate the
20   transcription unit of a gene so that the gene is turned off or silenced, and a corresponding protein product is not produced in a particular cell. For instance, one of the two X chromosomes in female mammals is inactivated or silenced by methylation.

DNA is extracted from a non-Mendelian disease sample using standard
25   techniques, known in the art, for isolating DNA from various samples such as cells , tissues, or organs, or other suitable specimens. Standard techniques for isolating DNA have are disclosed in reference textbooks or manuals such as Sambrook, Fritsch, and Maniatis, Molecular Cloning: A Laboratory Manual (1989), Cold Spring Harbor.

30

DNA may be extracted from any sample that may have epigenetic abnormalities associated with a non-Mendelian disease or any sample that exhibits characteristics of a non-Mendelian disease, for example, but not limited to cells of the following

- 24 -

tissues: Epithelial Tissues, Exocrine Glands, Endocrine Glands, Connective Tissues,
Adipose Tissue, Cartilage, Bone, Blood, Muscle Tissues comprising Smooth, Skeletal
or Cardiac Muscle Tissue, or Nervous Tissue comprising Brain Tissue.

5   Any methylation-sensitive restriction enzyme may be used for the purposes of
this invention. The terms "restriction endonucleases" and "restriction enzymes" refer
to bacterial enzymes, each of which cut double-stranded DNA at or near a specific
nucleotide sequence. The process of cutting or cleaving the DNA is referred to as
restriction digestion. The products of a restriction digestion are referred to as
10   restriction products. A restriction enzyme used in the present invention may yield
restriction products having blunt-ends or overhanging "sticky" ends. Specifically, a
restriction enzyme can symmetrically cut both strands of a double stranded DNA
fragment to produce a blunt-ended fragment, or a restriction enzyme may
assymetrically cleave the two strands of a DNA fragment to produce a DNA fragment
15   that has a single stranded overhang. In general, a methylation-sensitive restriction
enzyme used in the present invention will recognize and cleave a non-methylated
sequence, while it will not cleave a corresponding methylated sequence. Methylation
of plant and mammalian DNA occurs at CG or CNG sequences. This methylation
may interfere with the cleavage by some restriction endonucleases. Endonucleases
20   that are sensitive and not sensitive to m^5CG or m^5CNG methylation, as well as
isoschizomers of methylation-sensitive restriction endonucleases that recognize
identical sequences but differ in their sensitivity to methylation, can be extremely
useful for studying the level and distribution of methylation in eukaryotic DNA.
Examples of methylation-sensitive restriction enzymes, and corresponding restriction
25   site sequences, that can be used according to the present invention include, but are not
limited to: AatII (GACGTC); Bsh1236I (CGCG); Bsh1285I (CGRYCG); BshTI
(ACCGGT); Bsp68I (TCGCGA); Bsp119I (TTCGAA); Bsp143II (RGCGCY);
Bsu15I (ATCGAT); Cfr10I (RCCGGY); Cfr42I (CCGCGG); CpoI (CGGWCCG);
Eco47III (AGCGCT); Eco52I (CGGCCG); Eco72I (CACGTG); Eco105I
30   (TACGTA); EheI (GGCGCC); Esp3I (CGTCTC); FspAI (RTGCGCAY); HinlI
(GRCGYC); Hin6I (GCGC); HpaII (CCGG); Kpn2I (TCCGGA); MluI (ACGCGT);
NotI (GCGGCCGC); NsbI (TGCGCA); PauI (GCGCGC); PdiI (GCCGGC); Pfl23II

- 25 -

(CGTACG); Psp1406I (AACGTT); PvuI (CGATCG); SalI (GTCGAC); SmaI
(CCCGGG); SmuI (CCCGC); TaiI (ACGT); or TauI (GCSGC).

Size fractionation and purification of restricted DNA fragments can be
performed by any method known in the art, for example, but not limited to, separation
of DNA fragments of a desired size such as fragments of less than 10 kB by
centrifugation of a DNA fragment pool through a membrane or other suitable matrix
having size exclusion or inclusion properties. Alternatively, a pool of restricted DNA
fragments may be separated using agarose of polyacrylamide gel electrophoresis and
DNA fragments of a desired size may be purified using any suitable gel-extraction
composition such as glass milk or Quaternary ammonium ions. The desired size limit
of the fractionated and isolated DNA fragments depends on the size of the
endogenous DNA element that serves as a template for PCR amplification. As such
the "DNA fragments of a desired size" can be any size as long as they are larger than,
and can therefore comprise the endogenous DNA element.

As used, the terms "amplification," "amplify," or "amplifying," are defined as the
production of additional copies of a nucleic acid sequence and is generally carried out
using polymerase chain reaction (PCR) or other technologies well known in the art
(e.g., Dieffenbach and Dveksler, PCR Primer, a Laboratory Manual, Cold Spring
Harbor Press, Plainview NY [1995]). Nucleic acid amplification techniques allow for
increasing the concentration of a target or template sequence, or a portion or segment
thereof from a mixture of genomic DNA without cloning or purification. A review of
current nucleic acid amplification technology can be found in Kwoh et al., 8 Am.
Biotechnol. Lab. 14 (1990). In vitro nucleic acid amplification techniques include
polymerase chain reaction (PCR), transcription-based amplification system (TAS),
self-sustained sequence replication system (3SR), ligation amplification reaction
(LAR), ligase-based amplification system (LAS), Q.beta. RNA replication system and
run-off transcription. All present and future nucleic acid amplification technology can
be incorporated into the present invention. .

PCR is a preferred method for DNA amplification. PCR synthesis of DNA
fragments occurs by repeated cycles of heat denaturation of DNA fragments, primer

- 26 -

annealing onto endogenous sequence elements or exogenous adaptor ends of a DNA fragment or other suitable DNA template, and primer extension. These cycles can be performed manually or, preferably, automatically. Thermal cyclers such as the Perkin-Elmer Cetus cycler are specifically designed for automating the PCR process,

5   and are preferred. The number of cycles per round of synthesis can be varied from 2 to more than 50, and is readily determined by considering the source and amount of the nucleic

acid template, the desired yield and the procedure for detection of the synthesized DNA fragment.

10

        PCR techniques and many variations of PCR are known. Basic PCR techniques are described by Saiki et al. (1988 Science 239:487-491) and by K.B. Mullis in U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159, which are incorporated herein by reference.

15

        The conditions generally required for PCR include temperature, salt, cation, pH and related conditions needed for efficient amplification of at least a segment or portion of a DNA fragment template. PCR conditions include repeated cycles of heat denaturation, and incubation at a temperature permitting primer hybridization to an

20   endogenous sequence elements or exogenously ligated adaptors, and copying of the DNA fragment by the amplification enzyme. Heat stable amplification enzymes like the pwo, Thermus aquaticus or Thermococcus litoralis DNA polymerases are commercially available which eliminate the need to add enzyme after each denaturation cycle. The salt, cation, pH and related factors needed for enzymatic

25   amplification activity are available from commercial manufacturers of amplification enzymes.

        As provided herein an amplification enzyme is any enzyme which can be used for in vitro nucleic acid amplification, e.g. by the above-described procedures.

30   Amplification enzymes may be thermostable or thermolabile. Such amplification enzymes include pwo, Escherichia coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, T7 DNA polymerase, Thermus aquaticus (Taq) DNA polymerase, Thermococcus litoralis DNA polymerase, SP6

- 27 -

RNA polymerase, T7 RNA polymerase, T3 RNA polymerase, T4 polynucleotide kinase, Avian Myeloblastosis Virus reverse transcriptase, Moloney Murine Leukemia Virus reverse transcriptase, T4 DNA ligase, E. coli DNA ligase, Vent polymerases, or Q.beta. replicase. Preferred amplification enzymes are the pwo and Taq polymerases.

5   The pwo enzyme is especially preferred because of its fidelity in replicating DNA.

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by

10   avidin-enzyme conjugate detection; incorporation of 32P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

15

By the term "primer" is meant an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, capable of acting as a point of initiation of synthesis when placed under suitable conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand

20   is induced. Such suitable conditions comprise nucleotides and an amplification enzyme such as DNA polymerase and a suitable temperature, salt concentration, and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products.

25   The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, salt concentration , pH, source of primer and the use of the method. The primers of the present invention can hybridize or anneal to a sequence element that is endogenous to a DNA fragment template or the primers can

30   anneal to exogenous adaptor sequence elements that have been ligated to the ends of a DNA fragment template. Preferably, the primers anneal to an endogenous multi-copy DNA sequence element, for example, long or short interspersed nucleotide elements (LINEs or SINEs)..

Endogenous multi-copy DNA elements are repetitive DNA sequences that together are estimated to comprise 30% of total genomic sequences. Present at between $10 - 10^5$ copies per genome these multi-copy elements can be found

5   throughout the euchromatin and have been categorized as:

a) microsatellites / minisatellites (VNTR, DNA 'fingerprints)

b) dispersed-repetitive DNA, mainly transposable elements (LINES(for example, L1)/ SINES(foe example, Alu))

10   Endogenous multi-copy DNA elements can also include 'redundant' genes for histones, endogenous retroviral sequences (ERV), and ribosomal RNA and proteins, (gene-products present in cell in large numbers).

Many multi-copy DNA elements may be involved in regulation of gene

15   expression as they have been shown to be interspersed within single-copy sequences and have been shown to be located proximal to structural genes.

Long and short interspersed nucleotide elements (LINEs and SINEs), are represented in humans mainly by L1 (Furano AV. The biological properties and

20   evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol. 2000;64:255-94) and Alu elements (Watson et al., Molecular Biology of the Gene, fourth edition (1987) pp. 669-670), respectively. Both types of elements are considered to be retrotransposable (ie. can replicate via an RNA copy reinserted as DNA by reverse transcription) and they have significant roles in genomic function.

25   The inserted elements can be full length or truncated, or may be rearranged relative to full-length elements.

The most common and best characterised LINE is L1, having the following properties

30   Repeated approximately 50000 times in the human genome (0.5% of total)

Only about 3000 of these are full length; the remainder are truncated, mostly at the 5' end.

Full length element is about 6kb in size and contains two open reading frames, one of which encodes a reverse transcriptase.

AT-rich region is located near the 3' end of the element,

Element is flanked by two short direct repeats.

5

The main type of SINE is the Alu family, characterized as follows:

usually contain a target for the restriction enzyme Alu I;

$5 \times 10^5 - 10^6$ copies in the haploid genome, with an average of one repeat every 4 to 5 kb (1 - 10 % total);

10 Often present in the transcription unit of a gene, within introns and occasionally in non-translated regions of the mRNA;

Generally contain 300bp consensus sequence which consist of two tandem repeats of a 130bp sequence, one of which has a 32bp deletion, as such Alu family members are recognizably related in sequence, but not precisely conserved;

15 Elements are flanked by direct repeats;

Each repeat unit has an AT-rich region that suggests a poly A tail;

5' end resembles a pol III promoter region.

LINEs and SINEs both have a poly(A) tail which may act as a template for
reverse transcription from nicks made at the site of insertion in the host DNA by a
LINE-encoded endonuclease.

5

Primers of the present invention may be designed according to any L1 or Alu
sequence. For example, various analyses (Claverie,J.M. and Makalowski,W. Alu
alert, Nature 371, 752 (1994)) indicate that Alu repeats fall into 8 subfamilies, and
therefore, 8 ALU consensus sequences have been constituted and added to GenBank
10   as accession numbers U14567, U14568, U14569, U14570, U14571, U14572,
U14573 and U14574. A primer of the present invention may be designed in
accordance with any of these consensus sequences. For example, the deposited
consensus sequence of a subfamily of Alu repeats designated U14570 is as follows:
GGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGA
15   GGCGGGTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACAAG
G TGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGCGGTG
(SEQ ID NO:1)

Products of amplification reactions can be subjected to sequence
20   determinations. Amplification products, preferably PCR products, can optionally be
cloned into a vector before sequencing. When not cloning a PCR product, an adaptor
DNA elements can be ligated to the ends of PCR products, and the PCR products can
be sequenced using a primer that anneals to the adaptor element. Cloning, ligation,
and sequencing can be performed using standard techniques , such as protocols
25   described in textbooks or manuals such as Sambrook, Fritsch and Maniatis, Molecular
Cloning: A Laboratory Manual, 1989. Also, commercially available kits may be
utilized. Another alternative for sequence determination are automated DNA
sequencing systems and methods.

30   Nucleic acid sequences of amplification products isolated according to
methods of the present invention are disclosed in Figure 3. The region of the
chromosome to which a given sequence is located may be determined by
hybridization, including, but not limited to PCR amplification methods, or by
database searching.

Hybridization methods and conditions are well known in the art. Nucleic acids that are identical to the provided nucleic acid sequences, bind to the provided nucleic acid sequences (disclosed in Figure 3) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can determine a region of chromosome where a given sequence is located and thereby establish chromosomal loci for epigenetic abnormalities associated with a disease, including Mendelian or non-Mendelian disease.

Preferably, hybridization is performed using at least 15 contiguous nucleotides from any sequence identified by the methods of the present invention including, but not limited to, sequences disclosed in Figure 3. The probe will preferentially hybridize with a nucleic acid comprising a complementary sequence to the probe, allowing the identification of the chromosomal region of the nucleic acids of the biological material that uniquely hybridize to the selected probe. Probes of more than 15 nucleotides can be used, e.g. probes of from about 18 nucleotides up to the entire length of the provided nucleic acid sequences, but 15 nucleotides generally represents sufficient sequence for unique identification.

As mentioned above once the sequence (or a portion of the sequence) of a multi-copy DNA element has been isolated, this sequence can be used to map the location of the multi-copy DNA element on a chromosome. Accordingly, nucleic acids of the invention described herein or fragments thereof, can be used to map the location of multi-copy DNA elements of the invention on a chromosome. The mapping of the sequences of nucleic acids of the invention to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, sequences of the invention, for example, sequences disclosed in Figure 3, can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequences of nucleic acids of the invention. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human sequence

- 32 -

corresponding to the sequences of nucleic acids of the invention will yield an
amplified fragment.

5   Somatic cell hybrids are prepared by fusing somatic cells from different
mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow
and divide, they gradually lose human chromosomes in random order, but retain the
mouse chromosomes. By using media in which mouse cells cannot grow (because
they lack a particular enzyme), but in which human cells can, the one human
chromosome that contains the gene encoding a needed enzyme, depending on the
10   media, will be retained. By using various media, panels of hybrid cell lines can be
established. Each cell line in a panel contains either a single human chromosome or a
small number of human chromosomes, and a full set of mouse chromosomes,
allowing easy mapping of individual sequences to specific human chromosomes.
(D'Eustachio et al. (1983) Science 220:919-924). Somatic cell hybrids containing only
15   fragments of human chromosomes can also be produced by using human
chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a
particular sequence to a particular chromosome. Three or more sequences can be
20   assigned per day using a single thermal cycler. Using the sequences of nucleic acids
of the invention to design oligonucleotide primers, sublocalization can be achieved
with panels of fragments from specific chromosomes. Other mapping strategies which
can similarly be used to map a sequence of a nucleic acid of the invention to its
chromosome include in situ hybridization (described in Fan et al. (1990) Proc. Natl.
25   Acad. Sci. USA 87:6223-27), pre-screening with labeled flow-sorted chromosomes,
pre-selection by hybridization to chromosome specific cDNA libraries, and searching
of genomic databases.

Of course, persons skilled in the art will recognize that actual physical mapping of a
30   multi-copy DNA element on a chromosome, as described above, may not be
necessary where the multi-copy DNA element can be mapped in silico.

- 33 -

Once the sequence (or a portion of the sequence) of a multi-copy DNA element has been isolated, this sequence can be used to map the location of the gene on a chromosome by searching a genomic database, for example, but not limited to, a human genome database (www.genome.ucsc.edu/). Several genome databases are

5   also available from Celera Corp. or the National Center for Biotechnology Information (NCBI). Genome databases can be searched by comparing the known query sequence or reference sequence with genomic sequences stored and annotated in a database, and selecting sequences from the database that have a high similarity, preferably greater than 80% similarity, with the query or reference sequence.

10   Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 contiguous nucleotides long, more usually at least about 30 nucleotides long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the

15   art, such as BLAST, described in Altschul et al., J. Mol. Biol. (1990) 215:403-10.

To determine whether a nucleic acid exhibits similarity with the sequences presented herein, oligonucleotide alignment algorithms may be used, for example, but not limited to a BLAST (GenBank URL: www.ncbi.nlm.nih.gov/cgi-bin/BLAST/,

20   using default parameters: Program: blastn; Database: nr; Expect 10; filter: default; Alignment: pairwise; Query genetic Codes: Standard(1)), BLAST2 (EMBL URL: http://www.embl-heidelberg.de/Services/ index.html using default parameters: Matrix BLOSUM62; Filter: default, echofilter: on, Expect:10, cutoff: default; Strand: both; Descriptions: 50, Alignments: 50), or FASTA, search, using default parameters.

25

Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical, e.g., colcemid that disrupts the mitotic spindle.

30   The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases

- 34 -

have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic

5   Techniques (Pergamon Press, New York, 1988)). Sequences of isolated multi-copy DNA elements of the present invention that are shorter than 500 bases can be extended by any suitable technique, for example, a known sequence can be extended by a technique of genomic sequencing using a primer designed according to the known sequence.

10

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding

15   sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic

20   map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) Nature 325: 783-787.

25

Probes specific to the nucleic acids of the invention can be generated using a whole or portion of the nucleic acid sequences disclosed in Figure 3. The probes can be synthesized chemically or can be generated from longer nucleic acids using restriction enzymes. The probes can be labeled, for example, with a radioactive,

30   biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a nucleic acid of one of Figure 3. More preferably, probes are designed based on a contiguous sequence of one of the subject nucleic acids that remain unmasked following application of a masking program for masking low

- 35 -

complexity (e.g., XBLAST) to the sequence., i.e. one would select an unmasked region, as indicated by the nucleic acids outside the poly-n stretches of the masked sequence produced by the masking program. Probes are not only useful for determining chromosomal location of a sequence, but also can be used to determine
5   whether an epigenetic abnormality exists in another sample, for example a test sample obtained from a eukaryotic organism that exhibits symptoms of a disease, including Mendelian or non-Mendelian disease.

Once a chromosomal locus has been assigned to a multi-copy DNA element
10   obtained by the present invention, a genomic database or genetic map data can be used to identify one or more genes, for example about 1 to about 10 genes, that are proximal to the assigned chromosomal locus, preferably the identified one or more genes are physically adjacent to the assigned locus. Expression patterns of the genes in a Mendelian or non-Mendelian disease sample can then be compared against the
15   expression pattern of corresponding genes in a control sample to identify a gene having an epigenetically altered expression pattern. The disease sample and the control sample can be obtained from within the same organism, for example, without wishing to be limiting, expression of a gene within cancerous kidney cells could be compared against expression of a corresponding gene in a non-cancerous kidney cell
20   of the same organism. Alternately, the disease sample and the control sample can be obtained from different organisms. For example, without wishing to be limiting, expression of a gene in a prefrontal cortex sample from a schizophrenic individual can be compared against expression of a corresponding gene in a prefrontal cortex sample from a different non-schizophrenic individual. As another example, expression of a
25   gene in a cerebellum sample from a Huntingdon's disease patient can be compared against expression of a corresponding gene in a cerebellum sample obtained from a subject not suffering from Huntingdon's disease.

Techniques for determining expression patterns of genes are well known in the
30   art. For example, gene expression patterns oan be established using Northern analysis, reporter constructs such as GFP, quantitative PCR amplification, or DNA chip analysis (microarrays). If, for example, gene expression within a sample is determined using DNA chips, the mRNA from the sample is extracted, reverse

- 36 -

transcribed to the corresponding cDNA, amplified, fluorescently labeled and allowed

to hybridize with the sequences on a chip. Sequence-specific labels are captured on

the surface of the chip. By reading the fluorescence, one can determine which of the

genes were expressed and at what levels. DNA chip analysis is provided by several

5   companies, for example, but not limited to, Affymetrix and Nanogen. DNA chip

technology is an effective method for determining expression patterns of genes and

semiconductor fabrication technology has allowed for the packing of thousands of

gene sequences into square centimeter surfaces. Use of reporter constructs, Northern

analysis, and quantitative PCR amplification are equally effective alternatives.

10

Potential therapeutic approaches.

Detection of epigenetic abnormalities associated with diseases including, but

not limited to schizophrenia, diabetes, cancers, bipolar disorder, cystic fibrosis,

Duchennes muscular dystrophy, Huntington's disease and fragile X syndrome, may

15   lead to innovative DNA modification-based therapies. Recently a compound protein

consisting of a DNA methylation enzyme and a zinc-finger protein was constructed

(Xu G-L, Bestor TH. Nature Genetics 17: 376-379, 1997). The mechanism of action

of the protein consists of the recognition of a specific DNA sequence by the

zinc-finger protein that is specific for that sequence and subsequent modification of

20   the surrounding cytosines by DNA modification enzymes. A specific protein with

DNA modification enzyme restoring the normal pattern of DNA methylation can be

generated. The blood-brain barrier has been a major obstacle for the bloodborne

genetic constructs to reach the brain, but a recent study demonstrated that pegylated

neutral liposomes, unlike cationic ones, are stable in blood, do not get entrapped in

25   the lung, and are able to efficiently deliver plasmid DNA through the blood brain

barrier to the various sections of brain tissue .

The present invention provides methods and compositions for detecting DNA

elements that act as a marker for the specific dysfunctional genes and at the same time

30   identify the specific genes involved in diseases. Such information would lead quickly

to the development of a diagnostic test for such diseases, that could be incorporated

into a diagnostic kit. Further research on specific genes may also lead to treatment

options for people suffering from disease through either gene therapy work or through targeted drug development.

5   The heuristic value of epigenetics in diseases, including schizophrenia, derives from numerous important characteristics of epigenetic regulation of genes (Petronis A. Human morbid genetics revisited: relevance of epigenetics. Trends Genet. 2001 Mar;17(3):142-6). The epigenetic research program indicates that regulation of gene activity is critically important for normal functioning of the genome. Genes, even the ones that carry no mutations or disease predisposing polymorphisms, may be useless

10   or even harmful if not expressed in the appropriate amount, at the right time of the cell cycle, or in the right compartment of the nucleus. Epigenetic mechanisms, more so than DNA sequence-based ones, can explain a series of phenomenological features of a non-Mendelian disease, for example, in the case of, major psychosis including: i) relatively late age of onset and coincidence of the first symptoms with changes in the

15   hormonal status in the organism; ii) sexual dimorphism; iii) fluctuating course and sometimes recovery; iv) parental origin effects; and v) discordance of MZ twins. Furthermore, re-analysis of several etiological theories of major psychosis from an epigenetic point of view (Petronis A, Paterson AD, Kennedy JL. Schizophrenia: an epigenetic puzzle? Schizophrenia Bulletin 25:4: 639-655, 1999; Petronis A. The

20   genes for major psychosis: aberrant sequence or regulation? Neuropsychopharmacology, 23(1):1-12; 2000) suggested that epigenetic mechanisms have the potential to explain a number of clinical and molecular findings that traditionally have been supporting unrelated and somewhat antagonistic theories of schizophrenia and bipolar disorder, or have not been explained at all. Epigenetic

25   dysfunction may exhibit stability during meiosis and therefore can be transmitted from one generation to another (Klar AJ. Propagating epigenetic states through meiosis: where Mendel's gene is more than a DNA moiety. Trends Genet 1998; 14(8):299-301; Cavalli G, Paro R. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell 1998; 93(4):505-18;

30   Allen ND, Norris ML, Surani MA. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. Cell 1990 Jun 1;61(5):853-61; Silva AJ, White R. Inheritance of allelic blueprints for methylation patterns. Cell 1988 Jul 15;54(2):145-52; Morgan HD, Sutherland HG, Martin DI, and Whitelaw E (1999)

- 38 -

Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23: 314-8), which would simulate familial, i.e. genetic, cases of the disease.

5   The above description is not intended to limit the claimed invention in any manner, Furthermore, the discussed combination of features might not be absolutely necessary for the inventive solution.

The present invention will be further illustrated in the following examples. However, it is to be understood that these examples are for illustrative purposed only,
10   and should not be used to limit the scope of the present invention in any manner.

Examples

Example 1: Identification of loci having a hypomethylated sequence and a
15   retroelement in schizophrenia or bipolar disorder.

Brain tissues. Prefrontal cortex from post-mortem brains of individuals who were
20   affected with various psychiatric disorders (N=39; age at death [+S.D.] 40+12yr) and controls (N=9; age at death 48+7yr) were subjected to analysis. In the affected group, there were 26 males and 13 females, and the controls consisted of 8 males and 1 female. The distribution of psychiatric diagnoses was as follows: 11 bipolar disorder, 9 schizophrenia, 11 non-psychotic depression, and 8 psychosis NOS. The
25   overwhelming majority of the tested samples were from Caucasians, 1 American Black, and 2 Asians (all three affected). Brain tissues were kindly provided by the Stanley Foundation Brain Bank.

Methods. DNA samples were extracted from the brain tissues using a standard
30   phenol-chloroform extraction technique. Before the digestion of genomic DNA with a methylation sensitive restriction enzyme, an additional step of separation of the high molecular weight DNA (>15-20kb) from the partially degraded DNA was performed. The degraded DNA was removed by fractionation of 15 microgram of undigested

- 39 -

genomic DNA on a 1% low melting point agarose gel (Promega), cutting the agarose block that contained high molecular weight (>15-20kb) DNA, and incubating the block with an agarose- digesting enzyme, agarase, as recommended by the manufacturer (MBI Fermentas). After the agarose blocks were completely digested,

5   the high molecular weight DNA samples were digested with 50 units of methylation sensitive restriction enzyme, HpaII (MBI Fermentas) overnight. A test experiment using phage lambda DNA showed that the products of the agarase-treated agarose did not affect the ability of the restriction enzyme to cut DNA. In the next step, the unmethylated fraction of brain specific DNA was separated from the hypermethylated

10   fraction of DNA using a similar, gel-electrophoresis- based approach, during which DNA fragments smaller than arbitrarily selected 4 kb were cut out from the gel, purified using the NucleoSpin Extraction Kits (Clontech), and dissolved in 30 microliter of water. One to two microliter of the hypomethylated DNA solution were screened for the presence of·Alu sequences.

15

Alu sequences were sought using a protocol similat to the nested PCR protocol as in (Karlsson et al 2001) with primers that match the Alu sequences. Alu primer sequences were 'Alu For' GCCTGTACTCCCAGCAGTTT (SEQ ID NO:2) and 'Alu Rev' GGAGGGTGTTTGCACAATCT (SEQ ID NO:3). The reaction was performed

20   in 25 ul containing the standard PCR buffer, the two primers, 3 mM $MgCl_2$, 0.1 mM of dNTP, and 1U of Taq: Pfu polymerases mix (9:1). DNA template was denatured for 4 min at 94°C and amplification was performed in 30 cycles at 94°C, 58°C, and 72°C, 20 seconds each step. Alu PCR products were approximately 230 bp long.

25   PCR generated amplicons were cloned using the Qiagen PCR Cloningplus Kit. White E.coli colonies were grown up overnight, and plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen), and subjected to automated sequencing on the Perkin-Elmer/ABI 373A Sequencer (Automated DNA Sequencing Facility, York University, Toronto, Ontario).

30

The genomic location of the cloned sequences was identified using the UCSC Human Genome Project Working Draft, April 2002 assembly (http://genome.ucsc.edu/).

Table 1. The DNA samples that were selected for cloning and sequencing of individual *Alu*'s.

Sample #	Age	Sex	Ethnic background	Diagnosis
34	48	F	Caucasian	Bipolar Disorder
43	37	F	Caucasian	Bipolar Disorder
39	34	M	Caucasian	Mood disorder NOS
37	31	M	Caucasian	Schizophrenia
48	44	M	Caucasian	Schizophrenia
56	58	M	Caucasian	Schizophrenia
74	60	M	Caucasian	Schizophrenia
50	52	M	Caucasian	Control
57	44	M	Caucasian	Control

In the Alu amplification, however, agarose gel-visible (>0.1mg) PCR fragments were produced by about half of the DNA samples after 30 PCR cycles and nearly all samples if the number of cycles was increased to 35 or 40. Nine DNA samples (Table 1) that amplified the largest amount of Alu fragments were selected for further analysis, i.e. cloning and sequencing of individual Alu's. Ten to fifteen recombinant clones were sequenced from each PCR product, with a total of over 100 clones (some of these clones are presented in Fig. 4).

Genomic loci that exhibited higher than 95% of homology with the cloned Alu sequences were analyzed from two perspectives. In the first analysis, we investigated if Alu's mapped in the vicinity of known genes, and if so, how they could be related to abnormal brain functioning. The data of the Alu's mapping close to or within functional genes is presented in Table 2. About half of the Alu sequences (N=57) exhibited 100% sequence homology and mapped to Yq11.2, close to the testis transcript Y4. This indicates that the chromosome Y DNA contributed a significant portion of the hypomethylated DNA. The closest known gene to the Alu sequence on chromosome Y is the testis transcript Y4, the biological role of which is unknown. Other Alu sequences were scattered across the genome; their putative role in major psychosis is discussed in the next section.

Table 2. Cloned *Alu* sequences located within genes or in the close vicinity of genes

Clone Name	Homology length in bp; % Identity	Chr. Location	Gene Name
BD43 -A6-m	168bp; 100%	1q21	Protein kinase, AMP-activated, β2 (PRKAB2) (31Kb) KIAA1245 protein
BD43-RevE7m	191bp; 99.5%	1p31	Densin-180
BD34-A14M	187bp; 99%	2p23	Brain and reproductive organ-expressed gene (BRE) (TNFRSF1A modulator)*
BD43-E79m	186bp; 96.9%	2q37	Leucine rich repeat (in FLII) interacting (LRRFIP1)* Transcriptional repressor (GCF2)*
BD43-E78m BD43-E83m	192bp; 100%	5q22	U2 small nuclear ribonucleoprotein auxiliary (U2AF1RS1)
Sch56-m32	189bp; 99.5%	6p22.3	Ataxin 1 (SCA1)*
Sch37-m56	183bp, 96.5%	11q14.2	Embryonic ectoderm development protein WAIT-1
Sch74-E52m Sch74-E51m	192bp; 100%	17q12	AIOLOS isoform two (AIOLOS gene) (92Kb) KIAA1684 protein (6Kb)
Sch74-E318m	206bp; 97.7%	22q12	Oncostatin M (OSM)(5Kb) Leukemia inhibitory factor (LIF)(cholinergic) (25Kb) EBP50-PDZ interactor of 64 kD EP164 (19Kb) Splicing factor 3a, 120 kD SF3A1 (58Kb)

- 42 -

Numerous Sch and BD clones	191bp; 100%	Yq11	Testis transcript Y 4 (TTY4) (90Kb) HERV-K element (44Kb)
Ctrl57-E6m Ctrl50-RevE169m	187bp; 99% 179bp; 95%	1q31	Phosphatidylcholine 2-acylhydrolase (cPLA2)* Calcium-dependent phospholipid-binding protein (PLA2)
Ctrl50-E49m	185bp; 98%	2q36	Potassium voltage-gated channel, Isk-related KCNE4 (96Kb)
Ctrl57-E3m	191bp; 100%	5q34	WD repeat protein Gemin5* Mitochondrial ribosomal protein L22 MRPL22 (18Kb) CCR4-NOT transmission complex subunit 8 CNOT8 (60Kb)
Ctrl57-E5m	188bp; 99.0%	13q13	Lipoma HMGIC fusion partner LHFP (42Kb)
Numerous Ctrl clones	191bp; 100%	Yq11	Testis transcript Y4 (TTY4) (90Kb)

Clone ID consists of disease status (Sch - schizophrenia; BD - bipolar disorder; Ctrl -control), the number of the sample, and the clone number (following the hyphen). Asterisks indicate the *Alu* sequences that mapped within a gene. If *Alu* does not map within a gene, distance to the nearest known gene is indicated in brackets (kilobases; Kb)

5

The second analysis investigated if the cloned Alu sequences mapped to the genomic loci that showed evidence for linkage to SCZ and BD or revealed some chromosomal abnormalities (deletions, translocations) in individuals affected with major psychosis. The data of cloned Alu sequences that match the regions of putative linkage to major

10   psychosis are presented in Table 3. Since there is substantial overlap between the genetic loci predisposing to SCZ and the ones that increase the risk to BD (Berrettini 2000a; Berrettini 2000b; Cardno et al 2002), the type of psychosis - SCH or BD - was ignored in the matching of the cloned Alu's with the putatively linked genomic loci.

- 43 -

Table 3. Cloned *Alu* sequences that map to the regions of putative linkage to major psychosis

Clone Name	Homology length in bp; %Identity	Chr. Location	Evidence for linkage to schizophrenia or bipolar disorder (reference)
BD43-RevE77m	191bp; 99.5%	1p31	Rice et al 1997
BD43 -A6m	168bp; 100%	1q21	Brzustowicz et al 2000
BD43 -E78m	192bp; 100%	5q22	Straub et al 1997 Camp et al 2001 Bennett et al 1997[1]
Sch56-E32m	189bp; 99.5%	6p22	Kendler et al 2000 Schwab et al 1995a
Sch37-A9RR-m Sch56-E283m BD34-D19M BD34 -E62m	144bp; 99.4% 190bp; 99.5% 192bp; 100%	10p15 10p14	Straub et al 1998 DeLisi et al 2002 Faraone et al 1998 Schwab et al 1998
Sch56 -r-37m	186bp; 96.5%	11q14	Evans et al 1995; Petit et al 1999[2]
BD43 -15m	190bp; 99.5%	21q21	Detera-Wadleigh et al 1996
Sch74-E318_m Ctrl57-E4m	206bp; 97.7 % 193bp; 100 %	22q12.2	Pulver et al 1994 Gill et al 1996 Kelsoe et al 2001; Myles-Worsley et al 1999 Schizophrenia Collabporative Linkage Group

- 44 -

1998

Mujaheed et al 2000

DeLisi et al 2002; Moises et al 1995

Schwab et al 1995b

45 clones from affecteds and 12 clones from controls	191bp; 100 %	Yq11.2 Yq12	Alitalo et al 1988[3] Mors et al 2001[4]
Ctrl57-E6m Ctrl50-RevE169m	187bp; 99% 179bp; 95%	1q31.1	Detera-Wadleigh et al 1999
Ctrl57-E3m	191bp; 100 %	5q34	Crowe and Vieland 1999
Ctrl50-E166m	181bp; 100 %	18q23	Van Broeckhoven and Verheyen 1999; Verheyen et al 1999 Ewald et al 1999 Freimer et al 1996

1.   Interstitial deletion at 5q21-23.1 in an adult female with schizophrenia, mental retardation, and dysmorphic features.

2.   Schizophrenia-associated t(1;11)(q42.1;q14.3) breakpoint region.

3.   Translocation with the breakpoints between Yq11.23 and Yq12, and in 15p11, respectively, in two brothers who both had schizophrenia.

4.   The occurrence of the combined phenotype including both schizophrenia and bipolar disorder was significantly increased among individuals with the 47, XYY karyotype.

References of only positive findings of linkage to major psychosis are listed in the table.

Several of the genes listed within Table 2 are of significant interest, for example, the gene for spinocerebellar ataxia type 1 (SCA1)(6p22) (Tab. 2). SCA1 contains a

- 45 -

potentially unstable (CAG)n/(CTG)n trinucleotide repeat tract, which, when increased beyond the normal size, exhibits neurotoxic effects. In addition, the unstable trinucleotide repeats represent the molecular substrate for genetic anticipation, which, according to some authors (reviewed in (McInnis et al 1999)), is observed in major

5   psychosis. Some case-control and family-based association studies revealed statistically significant evidence that this gene is a predisposing factor to SCH (Joo et al 1999; Wang et al 1996).

Other genes listed in Table 2, although less known in the field of psychiatric research,

10   are also of significant interest. The embryonic ectoderm development gene (EED) (11q14) is necessary during gastrulation and organogenesis (Morin-Kensicki et al 2001). EED interacts with histone deacetylase (HDAC), a key player in the epigenetic regulation of chromatin structure, and the HDAC inhibitor trichostatin A, which relieves transcriptional repression mediated by EED (van der Vlag and Otte 1999).

15   Another link to the regulation of gene transcription can be found in a transcriptional repressor GCF2 (2q37), which exhibits differential affinity- depending on the DNA methylation status in that DNA methylation at the binding site abrogates both protein binding and repressor activity (Eden et al 2001).

20   The gene encoding leukemia inhibitory factor (LIF) (22q12) is expressed in the brain (Lemke et al 1997), promotes cholinergic expression in several neuronal populations (Cheema et al 1998), and plays a role in neuronal development, determination of phenotype, survival, and response to nerve injury (Moon et al 2002). Densin-180 (1p31) is highly concentrated at synapses along dendrites and it has been suggested

25   that this protein participates in specific adhesion between presynaptic and postsynaptic membranes at glutamatergic synapses. The mRNA encoding densin-180 is brain specific and is more abundant in forebrain than in cerebellum (Apperson et al 1996; Kennedy 1997). Four putative splice variants (A-D) of the cytosolic tail of densin-180 were shown to be differentially expressed during brain development

30   (Strack et al 2000). In this connection, it is interesting to note that one of the hypomethylated Alu sequences was found in the vicinity of the gene encoding splicing factor 3A (22q12) that is essential for the formation of the mature 17S U2 snRNP and the prespliceosome (Nesic and Kramer 2001). Alternative RNA splicing is

operating in a highly cell- and tissue-specific or developmentally specific manner. This directly applies to the neurons, where the functions of many gene products are regulated by alternative splicing (Shinozaki et al 1999). Differential splicing (e.g. mRNA for N-methyl-D-aspartate receptor (Le Corre et al 2000); dopamine D3

5   receptor (Karpa et al 2000)) has been implicated in SCH.

Several identified genes point at the putative immune and inflammatory components of major psychosis. Oncostatin M (OSM)(22q12) is a member of the interleukin (IL)-6 cytokine family that regulates inflammatory processes in the brain (Ruprecht et al

10   2001). Aiolos (17q12) encodes a hemopoietic-specific zinc finger transcription factor that is an important regulator of lymphocyte differentiation and is involved in the control of gene expression and, associated to nuclear complexes, participates in nucleosome remodeling (Schmitt et al 2002). It is not yet known if the gene encoding Aiolos can be expressed in the brain. A stress-responsive gene highly expressed in

15   brain and reproductive organs (BRE) (2p23) is a house-keeping gene that may play a role in homeostasis or in certain pathways of differentiation in cells of neural, epithelial, and germ line origins (Li et al 1995). Over expression of BRE inhibited TNF-induced NF kappa B activation, indicating that the interaction of BRE protein with the cytoplasmic region of p55 TNF receptor may modulate signal transduction

20   by TNF-alpha (Gu et al 1998).

Links to the metabolic stress in the affected brain is suggested by the gene encoding the AMP-activated protein kinase (beta 2 unit on chr 1q21). This kinase represents a heterotrimeric serine/threonine protein kinase with multiple isoforms for each subunit

25   (alpha, beta, and gamma) and is activated under conditions of metabolic stress. It is widely expressed in many tissues, including the brain (Turnley et al 1999).

Epigenetic studies of retroelements can be a valuable analytical (and diagnostic) tool that complements the more traditional genetic linkage, association, and gene

30   expression studies (Petronis et al 2000). Identification of the epigenetically dysregulated "junk" DNA sequences may allow for mapping of specific genomic regions in which genetic and/or epigenetic re-arrangements occurred. Such a retroelement may serve as a reporter, a signal that allows for the localization of

- 47 -

genomic changes, and a mechanism for the dysfunction of genes that are localized in such regions and may be the actual cause of psychosis. Expression studies of the genes located in the vicinity of epigenetic reporters can provide further clues to the pathobiological pathways of a disease. Of particular interest may be mapping of

5   differently regulated "junk" DNA elements performed in parallel with microarray-based global gene expression (Mirnics et al 2001). Large numbers of genes demonstrate differences in expression; however, it is never clear which changes are directly involved in the disease process and which ones just represent secondary 'downstream' changes and/or compensatory effects. There is no straightforward

10   approach for how to separate the two groups of events in the affected cell, but the presence of epigenetic changes in only some of the differentially expressed genes and the absence of such changes in the others can provide clues for a cause-effect relationship in the myriad of molecular changes in the affected brain. Support for this idea comes from the array-based studies in breast cancer, which detected numerous

15   differentially expressed genes in the malignant tissue and evident epigenetic deregulation of the otherwise impeccable BRCA1 (Hedenfalk et al 2001). Although the epigenetic status of other genes has not been investigated, hypermethylation of BRCA1 could certainly be one of the initiators of malignant growth.

20   Several Alu mapped loci have been of significant interest in linkage studies of major psychosis, including 1q21, 10p15, and 22q12, among numerous others (Table 3). Epigenetic mapping of hypomethylated retroelements may also facilitate genetic linkage studies. Traditional genetic linkage studies face major difficulties in fine mapping of the regions of susceptibility and identification of the actual gene

25   dysfunction that leads to major psychosis. Typically, the regions that exhibit evidence for linkage to major psychosis are in the range of ~10-15 mln nucleotides; furthermore, such regions may contain several hundred genes. Screening of such a large number of genes by traditional strategies for the detection of DNA variation is not a feasible task. Hypomethylated Alu's may pinpoint the very specific site of

30   genomic DNA and the critical gene(s) epigenetic dysfunction that may have caused psychosis. It is necessary to note that the putative epigenetic dysfunction may exhibit stability during meiosis and therefore can be transmitted from one generation to

- 48 -

another (Petronis 2001; Rakyan et al 2002), which would simulate familial cases of the disease.

5   Example 2: Identification of strong correlation between Huntingdon's Disease and hypomethylation in a locus having a retroelement.

10   *Brain tissues.* Samples from caudate and putamen (the brain regions that are primary sites of pathological changes in Huntington's disease [HD]) of HD patients (N=3; age at death 52+3 yr) and matched controls (n=4; age at death 54+3.5 yr) were analyzed.

Methods. Same as in Example 1 except for the following details. For the analysis of
15   Alu sequences within the Huntington's disease (HD) gene, primers for two Alu sequences downstream of the (CAG)n/(CTG)n trinucleotide repeat region were synthesized. It is of note that in the HD locus analysis, concrete Alu sequences were investigated, and the designed primers were complementary to the flanking regions of each specific Alu of the HD gene. This approach tested if DNA modification is
20   different in the regions surrounding Alu's within the gene that is known to cause a neuropsychiatric disease. The set of primers that amplified Alu located ~4Kb downstream of the (CAG)n/(CTG)n repeat region (NCBI ID: Z68756; Alu repeat region position 18,160bp -18,448bp) generated a visible PCR signal in the test experiments using genomic DNA as a template. This Alu was selected for further
25   analysis in the HD patients and controls. PCR conditions for amplification of this fragment were as follows: 1x standard PCR buffer, containing dimethylsulphoxide (DMSO) 10%; 2.5 mM $MgCl_2$; 0.16 mM dNTP and 10 microMolar of each of HD primer (1MF: CAGCGTACACATACACAGAAGAGA (SEQ ID NO:4) and 1MR: TTCCTAGTCACCAAGTCATAGCA (SEQ ID NO:5)), and 1U of Taq: Pfu
30   polymerases mix (9:1); 35 cycles at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec. PCR product size was ~360 bp.

- 49 -

The Alu sequence located ~4Kb downstream of the (CAG)n/(CTG)n repeat region of the HD gene was exclusively amplified in the hypomethylated fraction of the striatum DNA extracted from all three HD patients, but from none of the hypomethylated

5   fractions of the four controls. Thus, the striatum samples provided a 100% true positives and 0% false positives when diagnosing HD disease by identifying hypomethylation within a locus containing a retroelement. As such there is a strong correlation between HD disease and the identified locus.

10

The finding that HD Alu exhibited differential DNA methylation of the flanking regions in HD patients vs. controls supports the idea that epigenetic dysregulation of

15   retroelements sequences can lead to disease, for example neuropsychiatric diseases. This finding, suggests that analysis of differentially modified retroelements and their flanking sequences can point at the etiological disease genes.

It is interesting to note that HD represents a classical genetic disorder caused by

20   expansion of a (CAG)n/(CTG)n repeat tract. While epigenetic changes and their role in the disease have never been investigated in HD, there is indirect evidence that epigenetic factors may be operating in the regulation of the HD gene (Filippova et al 2001). The HD Alu data immediately link to our finding of an Alu within the gene for spinocerebellar ataxia type 1 (SCA1)(6p22) (see Example 1; Table 2). Like HD,

25   SCA1 contains a potentially unstable (CAG)n/(CTG)n trinucleotide repeat tract, which, when increased beyond the normal size, exhibits neurotoxic effects.

30   Example 3: Identification of strong correlation between Huntingdon's Disease and hypomethylation in a locus having a retroelement.

- 50 -

The same experiment as in Example 2 was repeated with 10 HD patients and 10
control subjects (see Table 4). DNA was extracted from cerebellum and striatum
samples for each HD patient and control subject.

5

Table 4. Data on Huntington Disease patients and control cases

Brain #	Distribution Dx	Age	Sex	PMI
B3976	H3	73	M	23.00
B4094	H3	72	M	12.75
B4381	H4	55	F	24.40
B5119	H3	68	F	17.00
B5146	H3	79	F	16.25
B5177	H3	49	M	25.25
B5331	Control	74	M	22.50
B5077	Control	67	M	18.50
B3813	Control	58	F	20.00
B5176	Control	65	F	24.25
B5113	Control	74	F	12.17
B5270	Control	52	M	22.56
B4781	H4	56	F	9.50
B4826	H4	49	M	16.60
B4828	H4	52	M	18.16
B5034	H4	54	M	20.08
B4739	Control	50	M	26.50
B4751	Control	54	M	24.20
B4974	Control	58	F	14.30
B5024	Control	56	M	21.33

Where H3 is the preterminal stage of HD

H4 is the terminal stage of HD

PMI is the postmortem interval (time between death and a brain tissue sampling)

The Alu sequence located ~4Kb downstream of the (CAG)n/(CTG)n repeat region of the HD gene was exclusively amplified in the hypomethylated fraction of the cerebellum DNA extracted from all 10 HD patients, but from none of the hypomethylated fractions of the 10 controls. Thus, the cerebellum samples provided a 100% correlation between HD disease and hypomethylation within a locus containing a retroelement.

With respect to striatum samples, the Alu sequence located ~4Kb downstream of the (CAG)n/(CTG)n repeat region of the HD gene was found to be amplified in the hypomethylated fraction of DNA from 8 out of 10 HD patients, and from only 1 out of 10 of the hypomethylated fractions of the four controls.

These results corroborate the findings and conclusions of Example 2. Persons skilled in the art will recognize that the methods provided in Examples 2 and 3 can be used for diagnosis of Huntingdon's disease, including pre-diagnosis of Huntingdon's disease.

Example 4: Detection of epigenetic abnormalities associated with schizophrenia or bipolar disorder.

Identification of the actual genes, which are epigenetically dysregulated and increase the risk to major psychosis, is not a simple task. Potentially any of the 35,000 human genes can be an epigenetic candidate for schizophrenia and bipolar

- 52 -

disorder. The present invention provides for epigenetic analysis of multicopy DNA sequences leading to the identification of DNA sequences that predispose to major psychosis. At least 35% of the human genome consists of numerous copies of different transposons dispersed in the genome (NB: only ~5% of the human genome

5   are exons, i.e. coding sequences of functional genes) (Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genetics, 13(8):335-40, 1997) . The range of copies of repetitive DNA fragments varies widely: There are 10^6 copies of Alu sequences and 10^5 copies L1 elements per genome (ibid.). The general opinion is that such sequences represent excess baggage

10   of our evolutionary heritage and do not perform any specific genomic function. This fraction of the genome is sometimes called "junk" or "parasitic" DNA. Such elements are not generally harmful to a cell as long as they do not exhibit any transcriptional activity and do not affect the integrity of the host genome. Transcriptional inactivation of the multicopy elements is achieved by their epigenetic modification. It has been

15   widely observed that DNA methylation plays a role in silencing various types of DNA sequences. Since it is becoming evident that DNA methylation may act in concert with histone acetylation (Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell, 88(4):471-81, 1997), chromatin conformation can also be considered a factor that plays a role in the

20   inactivation of retrotransposons as well as any other newly integrated DNA sequence. The findings that Alu and L1 elements as well as numerous other retroelements are methylated and transcriptionally inactive in the genomes of fungi, plants, and mammals provided the basis for postulating that epigenetic DNA modification represents a host genome defense system (Bestor TH. DNA methyltransferase in

25   genome defence. In: Epigenetic mechanisms of gene regulation. Eds: Russo VEA, Martienssen RA, Riggs AD. Cold Spring Harbor Laboratory Press, pp. 61-76, 1996; Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genetics, 13(8):335-40, 1997).

30   The epigenetic parameter may add a new dimension to the already available developments in psychiatric research. In our experiments we serendipitously detected that while the overwhelming majority of Alu sequences in the genomic DNA extracted from human brain are methylated, a small fraction of such sequences is

- 53 -

unmethylated. The origin of such selective Alu demethylation is not clear. Without wishing to be bound by theory, this most likely represents a local failure of the epigenetic host defense system, which has no direct impact to the normal functioning of the brain. On the other hand, such local epigenetic changes may not be limited to

5   the Alu sequences and may extend to the surrounding genes, causing dysregulation which may be detrimental to the cells. Supporting evidence for this comes from the observation that retroelements may become demethylated because they are located in the genomic region that was subjected to genetic and epigenetic re-organization. In malignant cells, it was detected that some Alu (Rubin CM, VandeVoort CA, Teplitz

10   RL, Schmid CW . Alu repeated DNAs are differentially methylated in primate germ cells. Nucleic Acids Research, 22(23):5121-7, 1994; Sinnett D, Richer C, Deragon JM, Labuda D. Alu RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. Journal of Molecular Biology, 226(3):689-706, 1992) and L1 (Florl AR, Franke KH, Niederacher D, Gerharz CD,

15   Seifert HH, Schulz WA. DNA methylation and the mechanisms of CDKN2A inactivation in transitional cell carcinoma of the urinary bladder. Laboratory Investigation, 80(10):1513-22, 2000; Jurgens B, Schmitz-Drager BJ, Schulz WA. Hypomethylation of L1 LINE sequences prevailing in human urothelial carcinoma. Cancer Research, 56(24):5698-703, 1996) elements became hypomethylated and

20   transcriptionally active.

        The present invention provides for identification of unmethylated "junk" DNA sequences in major psychosis allowing for mapping of specific genomic regions in which epigenetic re-arrangements occurred. Dysfunction of genes that are localized

25   such regions may be the actual cause of psychotic symptoms, while the demethylated multicopy element sequence would serve as a reporter, a signal that allows for localization of epigenetic changes in the genome.

        DNA samples were extracted from the frontal cortex of 40 post-mortem brain

30   tissues of individuals who were affected with schizophrenia and bipolar disorder as well as control individuals. In order to avoid artifacts related to partial brain DNA degradation (which may simulate hypomethylation and produce artifactual Alu amplification; see below), the following procedure was performed. Undigested total

- 54 -

genomic DNA was fractionated on an agarose gel, the high molecular weight
(>15-20kb) DNA was cut from the gel. The gel block, containing DNA, was treated
with a gel digesting enzyme, agarase. Without any additional procedures, such high
quality DNA samples can be further digested with a specific restriction enzyme and

5   subjected to further analyses. The methylation sensitive restriction enzyme, HpaII,
was used for digestion of DNA and the unmethylated fraction of brain specific DNA
(fragments smaller than arbitrarily selected 6kb) were separated from the methylated
fraction of DNA using gel electrophoresis. The <6kb fragments were purified from
the gel using glass milk. Screening for the presence of Alu's in the purified

10   unmethylated DNA was performed using PCR and primers complementary to the Alu
sequence. Alu amplicons were cloned into a vector and transformed into E.coli
XL1-blue. Up to ten recombinant clones from each PCR product were sequenced
from six individuals affected with major psychosis and four controls. The location of
such Alu sequences were identified using human genome databases

15   (http://genome.ucsc.edu/). It was detected that the Alu's from affected individuals in
numerous cases corresponded with the genomic regions that showed evidence for
linkage in genetic linkage studies of major psychosis. For example, one of the Alu
sequences cloned from an affected individual mapped to chr 1q21, the region that was
linked to schizophrenia (lod score of 6.5, the strongest evidence for linkage in

20   schizophrenia genetics thus far) in large multiplex schizophrenia families
(Brzustowicz LM, et al.,, 2000). In addition, an Alu clone from another psychosis
patient exhibited sequence homology with 1q42, the translocation region in a
schizophrenia kindred (St Clair D, et al. 1990). Other genomic regions where Alu
sequences mapped to the linkage 'spots', include 5q11 (although linkage to this region

25   [Sherrington R, et al.1988] was not replicated in other studies, two large kindreds
exhibit lod scores between 2 and 3 in favor of linkage). Other identified regions
include: 5q35 (chr 5 data reviewed in Crowe RR, et al. 1999), 8p23 (lod score 3.8 in a
large Swedish schizophrenia kindred), 8p21, 10p14, the pericentrometric regions of
chr 10 and 10q26 (Wildenauer DB, et. al. 1999), 11p15 and 11q13, 14q32 (Craddock

30   1999), 12p13 and 12q23-24 (Detera-Wadleigh SD. et al. 1999), and 22q13
(Nurnberger JI Jr, et al.1999). The 22q13 region exhibited evidence for linkage in
numerous studies and harbors a deletion region in velo-cardiofacial syndrome, a
disorder quite often resulting in psychotic symptoms (Chow EW, et al. 1994). For

- 55 -

more details on the localization of the cloned Alu sequences see Figure 1. Alu
sequences that are located in the vicinity (within 100,000 bp) of coding genes are
listed in Figure 2. Sequences of the cloned Alu's are provided in Figure 3.

5   The above results are of interest for the following reasons. First, clustering of
the Alu sequences into the groups of affected individuals and controls, if replicated in
an independent sample, would indicate that epigenetic changes of repetitive DNA
elements in some genomic loci are specific to major psychosis. This would be a
significant step forward in the light of the myriad of non-specific molecular changes
10   in the brains of patients affected with major psychosis. Second, genomic location of
the hypomethylated Alu's match with the loci that exhibit evidence for linkage to
major psychosis. Traditional genetic linkage studies face major difficulties in fine
mapping of the regions of susceptibility and identification of the actual gene
dysfunction that leads to major psychosis. Typically the regions that exhibit evidence
15   for linkage to major psychosis are in the range of ~10-40 cM, i.e. ~10-40 million
nucleotides (Thaker GK, et al., 2001; Tsuang MT, et al. 2001; Bray NJ, and Owen
MJ. 2001: Gershon ES. 2000; Nurnberger JI Jr, et al. 2000), and such regions contain
hundreds of genes. Screening of such a large number of genes by traditional strategies
for the detection of DNA variation is not possible. For fine mapping of prediposing
20   genes using the transmission disequilibrium test, very large samples are required; this
strategy has not been productive in psychiatric research thus far. In conclusion, the
"junk" DNA-based search for major psychosis genes may represent a valuable
'shortcut' in the identification of such genes. Hypomethylated Alu's may pinpoint very
specific sites of genomic DNA epigenetic dysfunction of which may cause major
25   psychosis.

Example 5: Identification of genes involved in etiology of schizophrenia or bipolar
disorder based on epigenetic analysis

30   The genes that are located in the regions exhibiting both linkage to major
psychosis and epigenetic abnormalities in Alu sequences are subjected to a detailed
analysis. Using the Celera Human Genome Database a list of genes from 1q21,
5q11, 8p23, 10p14, 11p15, 12p13, 12q23-24, 22q13, chr Y, and several other loci are

selected for further investigation from the epigenetic point of view. The list includes ~30 genes. Patients and controls are matched for age, sex, and race. Cases with drug and alcohol abuse are not used in the study. Treatment with neuroleptic medications is also a significant confounding factor. Neuroleptic naïve schizophrenic patients are

5   very rare, but cases with long neuroleptic free pre-mortem intervals are quite common. For example, in a recent study, one third of brain samples were neuroleptic-free for more than 6 months (Hernandez I, et al., 2000) and during this period, ~50% of schizophrenia patients are expected to relapse (Viguera AC, et al., 1997). Epigenetic dysregulation in schizophrenia and bipolar disorder, and other

10   disease associated epigenetic abnormalities in the brain may recur after neuroleptic treatment is stopped. Regarding the sample size, since there are no precedents of epigenetic studies in major psychosis, power analysis on the sample size is not possible. The investigation has been initiated with a relatively large sample by post-mortem brain study standards.

15

The prefrontal cortex from 25 post-mortem patients affected with major psychosis with >6 months of neuroleptic free period before death and a similar number of controls are used in the investigation. Over 70 brain samples from individuals who were affected with schizophrenia or bipolar disorder as well as controls are available

20   at our laboratory and this sample increases every year. Total mRNA from the brain tissues is extracted using standard RNA extraction techniques (Chomczynski P,et al., 1987) and subjected to reverse transcription and quantitative PCR amplification using the Bio-Rad Real Time PCR equipment (http://www.bio-rad.com/iCycler/). This experiment allows for the quantitative evaluation of the steady state level of the

25   candidate gene. 'Is it β-actin' mRNA serves as an internal standard for the degree of mRNA degradation. Expression of Is it β-actin is independent of the age of an individual and treatment (Schramm M, et al., 1999) and therefore can be reliably used as an estimate of the degree of post-mortem degradation. Steady state mRNA level of each individual gene is normalised according to its Is it β-actin mRNA data. The null

30   hypothesis is that the group of affected individuals exhibits no differences in the steady state mRNA levels of the selected genes in comparison to the group of controls. The genes that reject the null hypothesis, i.e. the ones that exhibit

statistically significant differences in steady state mRNA levels in affected tissues
versus controls, are subjected to further analysis. The problem is that not all genes that
exhibit significant differences in expression may carry epigenetic defects. Cases when
changes in steady state mRNA levels that may occur within hours or even minutes

5   after some triggers are applied, in the absence in any epigenetic changes in the
genome have to be excluded. Typically, epigenetic DNA modification targets
cytosines in CpG dinucleotides, each of which can be either methylated (metC) or
unmethylated (C). The gold standard technique for DNA methylation analysis is
based on the reaction of genomic DNA with sodium bisulfite under conditions such

10   that cytosine is deaminated to uracil but metC remains unreacted (Frommer M, et al.
1992). Sequencing of bisulfite modified DNA reveals which cytosines were
methylated and which cytosines were not. This approach has been fully
operationalized in our laboratory (Popendikyte V, et al., 1999). The present invention
provides for identifying one or more than one DNA coding sequences, from the list of

15   ~30 candidates, exhibiting disease specific epigenetic abnormality.

       All references are herein incorporated by reference.

20

- 58 -

REFERENCES

. Alitalo T, Tiihonen J, Hakola P, de la Chapelle A (1988): Molecular characterization of a Y;15 translocation segregating in a family. Hum Genet 79:29-35.

5 . Allen ND, Norris ML, Surani MA. Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. Cell 1990 Jun 1;61(5):853-61

. Apperson ML, Moon IS, Kennedy MB (1996): Characterization of densin-180, a new brain-specific synaptic protein of the O-sialoglycoprotein family. J Neurosci 16:6839-52.

10 . Bassett AS, Chow EW, Waterworth DM, Brzustowicz L (2001): Genetic insights into schizophrenia. Can J Psychiatry 46:131-7.

. Bennett RL, Karayiorgou M, Sobin CA, Norwood TH, Kay MA (1997): Identification of an interstitial deletion in an adult female with schizophrenia, mental retardation, and dysmorphic features: further support for a putative

15 schizophrenia-susceptibility locus at 5q21-23.1. Am J Hum Genet 61:1450-4.

. Berrettini W (2002): Review of bipolar molecular linkage and association studies. Curr Psychiatry Rep 4:124-9.

. Berrettini WH (2000a): Are schizophrenic and bipolar disorders related? A review of family and molecular studies. Biol Psychiatry 48:531-8.

20 . Berrettini WH (2000b): Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia. Biol Psychiatry 47:245-51.

. Bestor TH. DNA methyltransferase in genome defence. In: Epigenetic mechanisms of gene regulation. Eds: Russo VEA, Martienssen RA, Riggs AD. Cold Spring Harbor Laboratory Press, pp. 61-76, 1996.

25 . Bray NJ, Owen MJ. Searching for schizophrenia genes. Trends Mol Med. 2001; 7(4):169-74.

. Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS.Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22.Science 2000 Apr 28;288(5466):678-82

30 . Camp NJ, Neuhausen SL, Tiobech J, Polloi A, Coon H, Myles-Worsley M (2001): Genomewide multipoint linkage analysis of seven extended Palauan pedigrees with schizophrenia, by a Markov-chain Monte Carlo method. Am J Hum Genet 69:1278-89.

- 59 -

Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin P (2002): A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry 159:539-45.

Cavalli G, Paro R. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell 1998; 93(4):505-18

Cheema SS, Arumugam D, Murray SS, Bartlett PF (1998): Leukemia inhibitory factor maintains choline acetyltransferase expression in vivo. Neuroreport 9:363-6.

Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 1987 Apr;162(1):156-9.

Chow EW, Bassett AS, Weksberg R. Velo-cardio-facial syndrome and psychotic disorders: implications for psychiatric genetics. Am J Med Genet 1994;54(2):107-12

Craddock N, Lendon C. Chromosome Workshop: chromosomes 11, 14, and 15. Am J Med Genet. 1999 Jun 18;88(3):244-54.

Crowe RR, Vieland V. Report of the Chromosome 5 Workshop of the Sixth World Congress on Psychiatric Genetics. Am J Med Genet. 1999 Jun 18;88(3):229-32.

DeLisi LE, Shaw SH, Crow TJ, et al (2002): A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. Am J Psychiatry 159:803-12.

Detera-Wadleigh SD. Chromosomes 12 and 16 workshop. Am J Med Genet. 1999 Jun 18; 88(3):255-9.

Detera-Wadleigh SD, Badner JA, Berrettini WH, et al (1999): A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. Proc Natl Acad Sci U S A 96:5604-9.

Detera-Wadleigh SD, Badner JA, Goldin LR, et al (1996): Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q. Am J Hum Genet 58:1279-85.

Eden S, Constancia M, Hashimshony T, et al (2001): An upstream repressor element plays a role in Igf2 imprinting. Embo J 20:3518-25.

- 60 -

.   Ehrlich M and Ehrlich K (1993) Effect of DNA methylation and the binding of vertebrate and plant proteins to DNA. In: Jost JP and Saluz P (eds) DNA Methylation: Molecular Biology and Biological Significance pp. 145-168. Birkhauser Verlag, Basel, Switzerland.

5   .   Evans KL, Brown J, Shibasaki Y, et al (1995): A contiguous clone map over 3 Mb on the long arm of chromosome 11 across a balanced translocation associated with schizophrenia. Genomics 28:420-8.

.   Ewald H, Wang AG, Vang M, Mors O, Nyegaard M, Kruse TA (1999): A haplotype-based study of lithium responding patients with bipolar affective disorder

10   on the Faroe Islands. Psychiatr Genet 9:23-34.

.   Faraone SV, Matise T, Svrakic D, et al (1998): Genome scan of European-American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. Am J Med Genet 81:290-5.

.   Filippova GN, Thienes CP, Penn BH, et al (2001): CTCF-binding sites flank

15   CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet 28:335-43.

.   Florl AR, Franke KH, Niederacher D, Gerharz CD, Seifert HH, Schulz WA. DNA methylation and the mechanisms of CDKN2A inactivation in transitional cell carcinoma of the urinary bladder. Laboratory Investigation, 80(10):1513-22, 2000.

20   .   Freimer NB, Reus VI, Escamilla MA, et al (1996): Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet 12:436-41.

.   Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of

25   5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 1992;89:1827-31.

.   Gershon ES. Bipolar illness and schizophrenia as oligogenic diseases: implications for the future. Biol Psychiatry. 2000 Feb 1;47(3):240-4.

.   Gill M, Vallada H, Collier D, et al (1996): A combined analysis of D22S278

30   marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Schizophrenia Collaborative Linkage Group (Chromosome 22). Am J Med Genet 67:40-5.

- 61 -

.   Gonzalgo, M.L. and Jones, P.A. (1997) Mutagenic and epigenetic effects of DNA methylation. Mutat. Res. 386(2), 107-18

.   Gottesman II. Schizophrenia Genesis: The Origins of Madness. New York: W.H. Freeman; 1991.

5   .   Gu C, Castellino A, Chan JY, Chao MV (1998): BRE: a modulator of TNF-alpha action. Faseb J 12:1101-8.

.   Hedenfalk I, Duggan D, Chen Y, et al (2001): Gene-expression profiles in hereditary breast cancer. N Engl J Med 344:539-48.

.   Henikoff S, Matzke MA Exploring and explaining epigenetic effects. Trends
10 Genet 1997;13(8):293-5

.   Hernandez I, Sokolov BP. Abnormalities in 5-HT2A receptor mRNA expression in frontal cortex of chronic elderly schizophrenics with varying histories of neuroleptic treatment. J Neurosci Res. 2000; 59(2):218-25.

.   Johnston-Wilson NL, Sims CD, Hofmann JP, et al (2000): Disease-specific
15 alterations in frontal cortex brain proteins in schizophrenia, bipolar disorder, and major depressive disorder. The Stanley Neuropathology Consortium. Mol Psychiatry 5:142-9.

.   Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, and Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone
20 deacetylase to repress transcription. Nature Genetics 19: 187-91.

.   Joo EJ, Lee JH, Cannon TD, Price RA (1999): Possible association between schizophrenia and a CAG repeat polymorphism in the spinocerebellar ataxia type 1 (SCA1) gene on human chromosome 6p23. Psychiatr Genet 9:7-11.

.   .   Jurgens B, Schmitz-Drager BJ, Schulz WA. Hypomethylation of L1 LINE
25 sequences prevailing in human urothelial carcinoma. Cancer Research, 56(24):5698-703, 1996.

.   Karlsson H, Bachmann S, Schroder J, McArthur J, Torrey EF, Yolken RH (2001): Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci U S A 98:4634-9.

30 .   Karpa KD, Lin R, Kabbani N, Levenson R (2000): The dopamine D3 receptor interacts with itself and the truncated D3 splice variant d3nf: D3-D3nf interaction causes mislocalization of D3 receptors. Mol Pharmacol 58:677-83.

- 62 -

.   Kelsoe JR, Spence MA, Loetscher E, et al (2001): A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. Proc Natl Acad Sci U S A 98:585-90.

.   Kendler KS, Myers JM, O'Neill FA, et al (2000): Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry 157:402-8.

.   Kennedy MB (1997): The postsynaptic density at glutamatergic synapses. Trends Neurosci 20:264-8.

.   Klar AJ. Propagating epigenetic states through meiosis: where Mendel's gene is more than a DNA moiety. Trends Genet 1998; 14(8):299-301.

.   Lander E, Kruglyak L (1995): Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 11:241-247.

.   Le Corre S, Harper CG, Lopez P, Ward P, Catts S (2000): Increased levels of expression of an NMDARI splice variant in the superior temporal gyrus in schizophrenia. Neuroreport 11:983-6.

.   Lemke R, Gadient RA, Patterson PH, Bigl V, Schliebs R (1997): Leukemia inhibitory factor (LIF) mRNA-expressing neuronal subpopulations in adult rat basal forebrain. Neurosci Lett 229:69-71.

.   Li L, Yoo H, Becker FF, Ali-Osman F, Chan JY (1995): Identification of a brain- and reproductive-organs-specific gene responsive to DNA damage and retinoic acid. Biochem Biophys Res Commun 206:764-74.

.   Li TH, Kim C, Rubin CM, Schmid CW (2000): K562 cells implicate increased chromatin accessibility in Alu transcriptional activation. Nucleic Acids Res 28:3031-9.

.   Li TH, Schmid CW (2001): Differential stress induction of individual Alu loci: implications for transcription and retrotransposition. Gene 276:135-41.

.   Lyko, F. and Paro, R. (1999) Chromosomal elements conferring epigenetic inheritance. Bioessays 21(10), 824-32.

.   McInnis MG, McMahon FJ, Crow T, Ross CA, DeLisi LE (1999): Anticipation in schizophrenia: a review and reconsideration. Am J Med Genet 88:686-93.

.   McNeil TF (1995): Perinatal risk factors and schizophrenia: selective review and methodological concerns. Epidemiol Rev 17:107-12.

- 63 -

    .   Miniou P, Bourc'his D, Molina Gomes D, Jeanpierre M, Viegas-Pequignot E

(1997): Undermethylation of Alu sequences in ICF syndrome: molecular and in situ

analysis. Cytogenet Cell Genet 77:308-13.

    .   Mirnics K, Middleton FA, Lewis DA, Levitt P (2001): Analysis of complex

5   brain disorders with gene expression microarrays: schizophrenia as a disease of the

synapse. Trends Neurosci 24:479-86.

    .   Moises HW, Yang L, Li T, et al (1995): Potential linkage disequilibrium

between schizophrenia and locus D22S278 on the long arm of chromosome 22. Am J

Med Genet 60:465-7.

10 .   Moon C, Yoo JY, Matarazzo V, Sung YK, Kim EJ, Ronnett GV (2002):

Leukemia inhibitory factor inhibits neuronal terminal differentiation through STAT3

activation. Proc Natl Acad Sci U S A 99:9015-20.

    .   Morgan HD, Sutherland HG, Martin DI, and Whitelaw E (1999) Epigenetic

inheritance at the agouti locus in the mouse. Nature Genetics 23: 314-8.

15 .   Morin-Kensicki EM, Faust C, LaMantia C, Magnuson T (2001): Cell and

tissue requirements for the gene eed during mouse gastrulation and organogenesis.

Genesis 31:142-6.

    .   Mors O, Mortensen PB, Ewald H (2001): No evidence of increased risk for

schizophrenia or bipolar affective disorder in persons with aneuploidies of the sex

20 chromosomes. Psychol Med 31:425-30.

    .   Mowry BJ, Nancarrow DJ (2001): Molecular genetics of schizophrenia. Clin

Exp Pharmacol Physiol 28:66-9.

    .   Mujaheed M, Corbex M, Lichtenberg P, et al (2000): Evidence for linkage by

transmission disequilibrium test analysis of a chromosome 22 microsatellite marker

25 D22S278 and bipolar disorder in a Palestinian Arab population. Am J Med Genet

96:836-8.

    .   Myles-Worsley M, Coon H, McDowell J, et al (1999): Linkage of a composite

inhibitory phenotype to a chromosome 22q locus in eight Utah families. Am J Med

Genet 88:544-50.

30 .   Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with

abundant binding sites in genomic chromatin. Cell, 88(4):471-81, 1997.

.   Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, and Bird A (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386-9.

.   Nesic D, Kramer A (2001): Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. Mol Cell Biol 21:6406-17.

.   Nurnberger JI Jr, Foroud T. Chromosome 6 workshop report. Am J Med Genet. 1999 Jun 18;88(3):233-8.

.   Nurnberger JI Jr, Foroud T. Genetics of bipolar affective disorder. Curr Psychiatry Rep 2000 Apr;2(2):147-57.

.   Petit J, Boisseau P, Taine L, Gauthier B, Arveiler B (1999): A YAC contig encompassing the 11q14.3 breakpoint of a translocation associated with schizophrenia, and including the tyrosinase gene. Mamm Genome 10:649-52.

.   Petronis A. Human morbid genetics revisited: relevance of epigenetics. Trends Genet. 2001 Mar;17(3):142-6.

.   Petronis A, Paterson AD, Kennedy JL. Schizophrenia: an epigenetic puzzle? Schizophrenia Bulletin 25:4: 639-655, 1999

.   Petronis A. The genes for major psychosis: aberrant sequence or regulation? Neuropsychopharmacology, 23(1):1-12; 2000.

.   Petronis A, Gottesman, II, Crow TJ, et al (2000): Psychiatric epigenetics: a new focus for the new century. Mol Psychiatry 5:342-6.

.   Popendikyte V, Laurinavicius A, Paterson AD, Macciardi F, Kennedy JL, Petronis A. DNA methylation at the putative promoter region of the human dopamine D2 receptor gene. Neuroreport 1999;10:1249-55.

.   Pulver AE, Karayiorgou M, Wolyniec PS, et al (1994): Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: Part 1. Am J Med Genet 54:36-43.

.   Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E (2002): Metastable epialleles in mammals. Trends Genet 18:348-51.

.   Razin, A. and Shemer, R. (1999) Epigenetic control of gene expression. Results Probl. Cell. Differ. 25, 189-204

- 65 -

.   Rice JP, Goate A, Williams JT, et al (1997): Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 1, 6, 8, 10, and 12. Am J Med Genet 74:247-53.

.   Riggs A, Xiong Z, Wang L, and LeBon JM (1998) Methylation dynamics, epigenetic fidelity and X chromosome structure. In: Wolffe AP (ed) Epigenetics, pp. 214-227. John Wiley & Sons, Chistester.

.   Robertson KD and Wolffe AP (2000) DNA methylation in health and disease. Nature Review Genet 1:11-9

.   Rubin CM, VandeVoort CA, Teplitz RL, Schmid CW . Alu repeated DNAs are differentially methylated in primate germ cells. Nucleic Acids Research, 22(23):5121-7, 1994.

.   Ruprecht K, Kuhlmann T, Seif F, et al (2001): Effects of oncostatin M on human cerebral endothelial cells and expression in inflammatory brain lesions. J Neuropathol Exp Neurol 60:1087-98.

.   Schizophrenia Collaborative Linkage Group (1998): A transmission disequilibrium and linkage analysis of D22S278 marker alleles in 574 families: further support for a susceptibility locus for schizophrenia at 22q12. Schizophr Res 32:115-21.

.   Schmitt C, Tonnelle C, Dalloul A, Chabannon C, Debre P, Rebollo A (2002): Aiolos and Ikaros: Regulators of lymphocyte development, homeostasis and lymphoproliferation. Apoptosis 7:277-84.

.   Schramm M, Falkai P, Tepest R, Schneider-Axmann T, Przkora R, Waha A, Pietsch T, Bonte W, Bayer TA. Stability of RNA transcripts in post-mortem psychiatric brains. J Neural Transm. 1999;106(3-4):329-35.

.   Schwab SG, Albus M, Hallmayer J, et al (1995a): Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nat Genet 11:325-7.

.   Schwab SG, Hallmayer J, Albus M, et al (1998): Further evidence for a susceptibility locus on chromosome 10p14-p11 in 72 families with schizophrenia by nonparametric linkage analysis. Am J Med Genet 81:302-7.

.   Schwab SG, Lerer B, Albus M, et al (1995b): Potential linkage for schizophrenia on chromosome 22q12-q13: a replication study. Am J Med Genet 60:436-43.

- 66 -

. Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H. Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature. 1988 Nov 10;336(6195):164-7.

5 . Shinozaki A, Arahata K, Tsukahara T (1999): Changes in pre-mRNA splicing factors during neural differentiation in P19 embryonal carcinoma cells. Int J Biochem Cell Biol 31:1279-87.

. Shinozaki A, Arahata K, Tsukahara T (1999): Changes in pre-mRNA splicing factors during neural differentiation in P19 embryonal carcinoma cells. Int J Biochem Cell Biol 31:1279-87.

10 . Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. Nat Genet 1999 Jun;22(2):203-206

. Silva AJ, White R. Inheritance of allelic blueprints for methylation patterns. Cell 1988 Jul 15;54(2):145-52

. Sinnett D, Richer C, Deragon JM, Labuda D. Alu RNA transcripts in human
15 embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. Journal of Molecular Biology, 226(3):689-706, 1992.

. St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, Gosden C, Evans HJ. Association within a family of a balanced autosomal translocation with major mental illness. Lancet. 1990 Jul 7;336(8706):13-6.

20 . Strack S, Robison AJ, Bass MA, Colbran RJ (2000): Association of calcium/calmodulin-dependent kinase II with developmentally regulated splice variants of the postsynaptic density protein densin-180. J Biol Chem 275:25061-4.

. Straub RE, MacLean CJ, Martin RB, et al (1998): A schizophrenia locus may be located in region 10p15-p11. Am J Med Genet 81:296-301.

25 . Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS (1997): Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families. Mol Psychiatry 2:148-55.

. Susser E, Neugebauer R, Hoek HW, et al (1996): Schizophrenia after prenatal famine. Further evidence. Arch Gen Psychiatry 53:25-31.

30 . Thaker GK, Carpenter WT Jr. Advances in schizophrenia. Nat Med. 2001 Jun;7(6):667-71.

. Tsuang MT, Stone WS, Faraone SV. Genes, environment and schizophrenia. Br J Psychiatry Supl. 2001 Apr;40:s18-24.

- 67 -

. Turnley AM, Stapleton D, Mann RJ, Witters LA, Kemp BE, Bartlett PF (1999): Cellular distribution and developmental expression of AMP-activated protein kinase isoforms in mouse central nervous system. J Neurochem 72:1707-16.

. Van Broeckhoven C, Verheyen G (1999): Report of the chromosome 18 workshop. Am J Med Genet 88:263-70.

. van der Vlag J, Otte AP (1999): Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. Nat Genet 23:474-8.

. Verdoux H, Geddes JR, Takei N, et al (1997): Obstetric complications and age at onset in schizophrenia: an international collaborative meta-analysis of individual patient data. Am J Psychiatry 154:1220-7.

. Verheyen GR, Villafuerte SM, Del-Favero J, et al (1999): Genetic refinement and physical mapping of a chromosome 18q candidate region for bipolar disorder. Eur J Hum Genet 7:427-34.

. Viguera AC, Baldessarini RJ, Hegarty JD, van Kammen DP, Tohen M. Clinical risk following abrupt and gradual withdrawal of maintenance neuroleptic treatment.Arch Gen Psychiatry 1997; 54(1):49-55

. Wang S, Detera-Wadleigh SD, Coon H, et al (1996): Evidence of linkage disequilibrium between schizophrenia and the SCa1 CAG repeat on chromosome 6p23. Am J Hum Genet 59:731-6.

. Wildenauer DB, Schwab SG. Chromosomes 8 and 10 workshop. Am J Med Genet. 1999 Jun 18;88(3):239-43.

. Xu G-L, Bestor TH. Nature Genetics 17: 376-379, 1997.

. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genetics, 13(8):335-40, 1997.

The present invention has been described with regard to preferred embodiments. However, it will be obvious to persons skilled in the art that a number of variations and modifications can be made without departing from the scope of the invention as described herein.